DOCUMENT RESUME

ED 214 783 SE 036 474

AUTHOR Schey, Harry M.; And Others

TITLE UMAP Modules-Units 60-62, 67, 68, 70.
INSTITUTION Education Development Center, Inc., Newton, Mass.
SPONS AGENCY National Science Foundation, Washington, D.C.
PUB DATE 80
GRANT SED76-19615; SED76-19615-A02
NOTE 136p.; Contains occasional light type.

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Biological Sciences; *Calculus; *College Mathematics; Economics Education; Genetics; Higher Education; Instructional Materials; *Learning Modules; *Mathematical Applications; Mathematical Enrichment; Mathematical Models; Population Growth; *Problem Solving; *Supplementary Reading Materials
IDENTIFIERS Nervous System

ABSTRACT

The first three units in this set of mathematical applications cover applications of calculus to economics. Under the title, The Distribution of Resources, are the following: 60-The Lorenz Curve; 61-The Equal Share Coefficient; and 62-The GINI Index. Each of these units contains a set of exercises, a model exam, and answers to both the exercises and exam. The second three units cover biological applications of calculus. This includes: 67-Modeling the Nervous System-Reaction Time and the Central Nervous System; 68-Population Growth and the Logistic Curve; and 70-Selection in Genetics. All three contain problem exercises and answers to at least some problems. All six of the units are intended as supplementary enrichment material for undergraduate students. (MP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED214783

SE036474

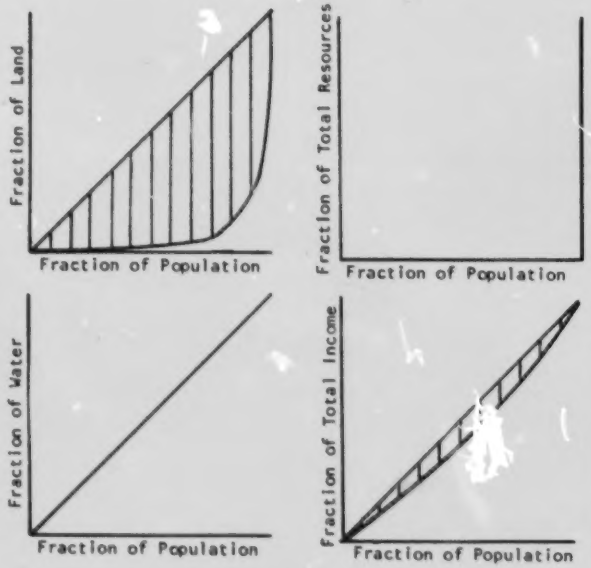
umap

UNITS 60, 61, 62

MODULES AND MONOGRAPHS IN UNDERGRADUATE MATHEMATICS AND ITS APPLICATIONS PROJECT

THE DISTRIBUTION OF RESOURCES

by Harry M. Schey



APPLICATIONS OF CALCULUS TO ECONOMICS

edc/umap /55chapel st./newton, mass. 02160

THE DISTRIBUTION OF RESOURCES

by

Harry M. Schey
Department of Biostatistics
University of North Carolina
Chapel Hill North Carolina 27514

TABLE OF CONTENTS

INTRODUCTION 1

1. THE LORENZ CURVE (U 60) 2

1.1 Distribution Tables and Their Associated Graphs 2
1.2 Complete Equality 5
1.3 Complete Inequality 8
1.4 Exercises 11
1.5 Model Exam 14
1.6 Answers to Exercises 17
1.7 Answers to Model Exam 21

2. THE EQUAL SHARE COEFFICIENT (U 61) 25

2.1 Introduction 25
2.2 The Slope of a Lorenz Curve 25
2.3 The Definition 29
2.4 A Geometric Approach 30
2.5 A Numerical Approach 33
2.6 Linear Interpolation 36
2.7 Exercises 37
2.8 Model Exam 39
2.9 Answers to Exercises 42
2.10 Answers to Model Exam 47

3. THE GINI INDEX (U 62) 49

3.1 Introduction 49
3.2 Area Under the Lorenz Curve 49
3.3 The Definition 51
3.4 Using the Definite Integral 53
3.5 The Rectangle Method 54
3.6 A Numerical Method 58
3.7 Exercises 61
3.8 Model Exam 62
3.9 Answers to Exercises 65
3.10 Answers to Model Exam 70

Intermodular Description Sheet: UMAP Units 60-62

Title: THE DISTRIBUTION OF RESOURCES

Author: Harry M. Schey
Department of Biostatistics
University of North Carolina
Chapel Hill, NC 27514

Review Stage/Date: IV 4/6/79

Classification: ECON APPL CALC/DIS RESOURCES

Prerequisite Skills:

Unit 60:
1. Be able to read and draw graphs (cartisian coordinates only).

Unit 61:
1. Understand the relationship between a Lorenz curve and the equality of a resource distribution.
2. Understand the concepts of slope and derivative and the relation between them.

Unit 62:
1. Understand the concept of the definite integral.

Output Skills:

Unit 60:
1. Define and describe what is meant by a Lorenz curve.
2. Read and interpret a Lorenz curve.
3. Draw a Lorenz curve from given data.
4. Compare the equality of two or more distributions on the basis of their Lorenz curves.
5. Draw the Lorenz curves corresponding to complete equality and complete inequality.

Unit 61:
1. Define the equal share coefficient.
2. Describe the significance of the slope of a Lorenz curve.
3. Show how the equal share coefficient is a measure of the equality of a resource distribution.
4. Be able to calculate equal share coefficients, graphically, numerically, and by differentiation.

Unit 62:
1. Define the Gini index.
2. Show how the Gini index is a measure of the inequality of a resource distribution.
3. De able to calculate the Gini index of a given resource distribution.

Other Related Units:

© 1979 EDC/Project UMAP
All rights reserved.

MODULES AND MONOGRAPHS IN UNDERGRADUATE

MATHEMATICS AND ITS APPLICATIONS PROJECT (UMAP)

The goal of UMAP is to develop, through a community of users and developers, a system of instructional modules in undergraduate mathematics and its applications which may be used to supplement existing courses and from which complete courses may eventually be built.

The Project is guided by a National Steering Committee of mathematicians, scientists and educators. UMAP is funded by a grant from the National Science Foundation to Education Development Center, Inc., a publicly supported, nonprofit corporation engaged in educational research in the U.S. and abroad.

PROJECT STAFF

Ross L. Finney	Director
Solomon Garfunkel	Associate Director/Consortium Coordinator
Felicia DeMay	Associate Director for Administration
Barbara Kelczewski	Coordinator for Materials Production
Edwina Michener	Editorial Consultant
Dianne Lally	Project Secretary
Paula M. Santillo	Administrative Assistant
Carol Forray	Production Assistant
Zachary Zevitas	Order Processor

NATIONAL STEERING COMMITTEE

W.T. Martin	MIT (Chairman)
Steven J. Brams	New York University
Llayron Clarkson	Texas Southern University
Ernest J. Henley	University of Houston
Donald A. Larson	SUNY at Buffalo
William F. Lucas	Cornell University
R. Duncan Luce	Harvard University
George Miller	Nassau Community College
Frederick Mosteller	Harvard University
Walter E. Sears	University of Michigan Press
George Springer	Indiana University
Arnold A. Strassenburg	SUNY at Stony Brook
Alfred B. Willcox	Mathematical Association of America

This material was prepared with the support of National Science Foundation Grant No. SED76-19615 A02. Recommendations expressed are those of the author and do not necessarily reflect the views of the NSF, nor of the National Steering Committee.

THE DISTRIBUTION OF RESOURCES

INTRODUCTION

The people and the nations of the world own many resources; money, land, energy, food, oil, etc. One of the most important questions about these resources has to do with how they are distributed among people or among groups of people. In this module we investigate several ways in which mathematics can be used to measure the distribution of resources.

1. THE LORENZ CURVE

1.1 Distribution Tables and Their Associated Graphs

Table I shows how total income was distributed among Americans in 1955.* It shows that the lowest .2 of the people (that is, the 20% of the population having the lowest annual income) received only .05 of the total income, the lowest .4 (which includes the lowest .2) received .16 of the total income, etc. This table thus gives us the kind of information we'll work with in this module, since it tells us how a certain resource (income) is distributed among a certain group of people (Americans in 1955). But often a table of numbers does not convey information as well as a picture or a diagram. As we'll see now, a graph of the numbers in Table I will give us a valuable way to visualize the distribution of income.

Table I

Fraction of People	Fraction of Income
0.0	0.0
.2	.05
.4	.16
.6	.33
.8	.55
1.0	1.00

The graph shown in Figure 1 is made by drawing a smooth curve through the points that represent the numbers in Table I. Note that we use an ordinary Cartesian

*Samuelson, P.A., Economics, an Introductory Analysis, 4th ed. (Mc.Graw-Hill, New York) 1958, page 69.

coordinate system. The fraction of the population is plotted on the horizontal axis, and the fraction of total income, on the vertical. Since neither of these fractions is ever negative, we need only the first quadrant for the graph. (In fact, since the numbers involved never exceed 1, we only need the unit square.)

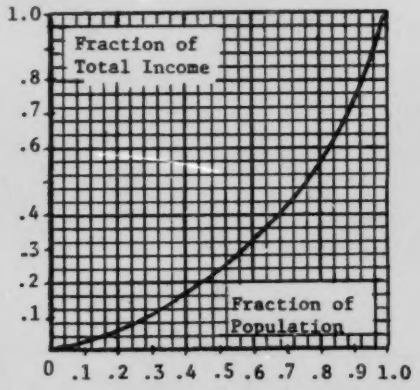


Figure 1.

A graph such as Figure 1 shows how some resource is distributed. Such a monotonic, concave curve is called a *Lorenz curve*. Note that a Lorenz curve always passes through the points (0,0) and (1,1).

It should be emphasized that to draw this Lorenz curve we plotted the points given in Table I and then *assumed* that those points can be connected by a smooth curve. In doing this we "create" values that do not appear in Table I. Although the fraction of total income corresponding to, say, 0.35 of the population may in reality not be exactly the value given by our curve, our assumption is that it will not be very different.

Exercise 1. Use the curve drawn in Figure 1 to fill in Table II.

Table II

Fraction of People	Fraction of Income
0.1	
.35	
.5	
.75	
.9	

Exercise 2.

Table III shows how income was distributed in the United States in 1971.* Draw the Lorenz curve for this distribution.

Table III

Fraction of People	Fraction of Income
0.0	0.0
.2	.06
.4	.18
.6	.36
.8	.60
1.0	1.00

*Samuelson, op. cit., 9th ed., 1973, page 85.

The Lorenz curve in Figure 1 does give a better "picture" of income distribution than the bare numbers in Table I, but it still does not provide any way to decide whether income was divided equally among Americans in 1955 or whether it was divided in a highly unequal way. What we need is some basis for comparison so that we can look at a Lorenz curve and judge how close to (or far from) equality the distribution is. To obtain this basis, we examine the two extreme cases: complete equality and complete inequality.

1.2 Complete Equality

What do we mean by complete equality? Table IV shows how income is distributed in the case of complete equality: The lowest .2 of the population (which in this case is *any* .2) gets .2 of the income, the lowest .4 gets .4 of the income, etc. In Exercise 3, we investigate the Lorenz curve for this *completely equal* distribution.

Table IV

Fraction of People	Fraction of Income
0.0	0.0
.2	.2
.4	.4
.6	.6
.8	.8
1.0	1.0

Exercise 3. In Figure 2, we have plotted two of the points from Table IV. Plot the remaining points in Figure 2. Connect the points to verify that in this case of complete equality, the Lorenz "curve" is a straight

line which makes an angle of 45° with the horizontal axis as shown in Figure 3.

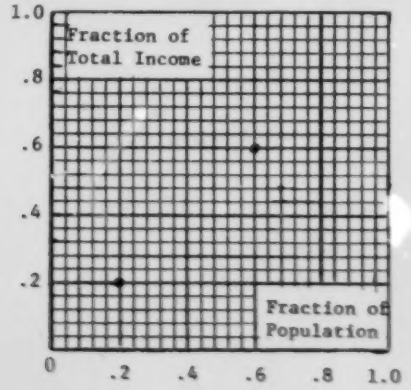


Figure 2.

Figure 3 thus shows a Lorenz curve for the case of perfect equality. Notice that the vertical axis is labeled "Fraction of Resource" instead of "Fraction of Total Income" to emphasize that this is the form of a Lorenz curve in the case of perfect equality in the distribution of *any* resource, not just total income. A Lorenz curve like the one you drew in Figure 2 and the one shown in Figure 3 is called the "curve of absolute equality." (See Figure 4).

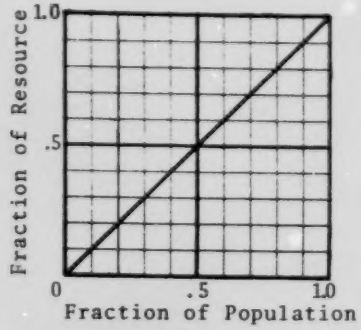


Figure 3.

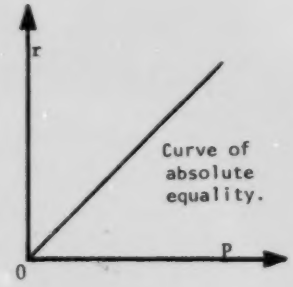


Figure 4.

Exercise 4. Let the fraction of population be denoted p and the fraction of the resource be denoted r. Use these two symbols to write an equation for the curve of absolute equality (Figure 4).

Go back to the Lorenz curve with which we began, the one shown in Figure 1. It is redrawn in Figure 5, but as you can see, something new has been added: we have also shown the curve of absolute equality, which is how the Lorenz curve *would* look *if* income were distributed perfectly equally. So now, at last, we have the beginnings of a basis for comparisons: A realistic Lorenz curve "sags" below the curve of absolute equality, and the more it sags, the greater the inequality of the distribution it represents.

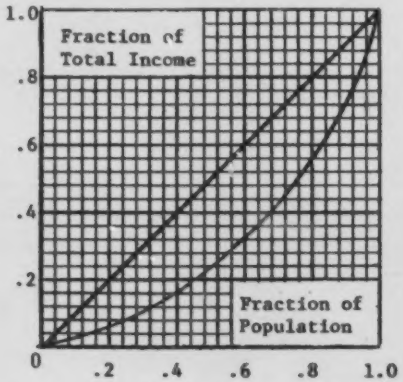


Figure 5.

Exercise 5. Table V shows how arable land is distributed among farmers in Bolivia, Denmark, and the United States.* Draw the Lorenz curve for each of the three countries. In which country is the land most equally distributed among the farmers? In which country is it the least? Do these curves suggest a way to measure inequality? (See Sections 2 and 3.)

*Based on data from B. Russett, World Handbook of Social and Political Indicators, 1964. As quoted by Harf, Ledford, and Thomson, Learning Packages in Data Analysis for International Studies, No. 4, Behavioral Sciences Laboratory, The Ohio State University, 1974.

Table V

Fraction of Farmers	Fraction of Land (1964)		
	Bolivia (x)	Denmark (•)	United States (o)
0.0	0.0	0.0	0.0
.1	.0	.06	.025
.2	.0	.12	.05
.3	.0	.18	.075
.4	.0	.24	.10
.5	.010	.30	.13
.6	.016	.36	.18
.7	.022	.45	.22
.8	.03	.54	.28
.9	.04	.70	.42
1.0	1.00	1.00	1.00

1.3 Complete Inequality

Next we discuss the other extreme case: complete inequality. We already know that a Lorenz curve which represents an unequal distribution of some resource sags below the curve of absolute equality as in Figure 5. The question we'll now examine is "How much can a Lorenz curve sag?"

The answer is that the Lorenz curve which sags as far as possible is one which represents a completely unequal distribution. What, then, is a completely unequal distribution? One example is the situation in which almost no one has any income; the exception is the one person who gets the entire amount. That distribution is shown in Table VI. Notice that the lowest .2, .4, ..., .9, .99, .999, ... of the population gets nothing. The reason the last entry

in Table VI is a 1 instead of a 0 is that *all* the people get *all* the income even though it is divided in such a way that it all goes to one person.

Table VI

Fraction of People	Fraction of Income
0.0	0.0
0.2	0.0
0.4	0.0
0.6	0.0
0.8	0.0
1.0	1.0

The Lorenz curve for the distribution in Figure 6a shows the completely unequal distribution recorded in Table VI. Its graph is very discontinuous at 1, since it "jumps" from 0 to 1. When we smoothly connect the data points as in Figure 6b, the curve has a very severe bend or "corner" at the point p = 1.

Now, even though a well-defined function would not have a vertical line segment in its graph, the curve suggested to us can be idealized as having the right angle shown in Figure 6c: this Lorenz curve lies along the p-axis almost all the way to p = 1 and then "shoots" up vertically.

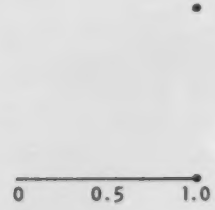


Figure 6a.

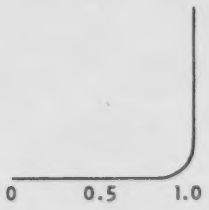


Figure 6b.

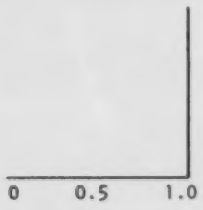


Figure 6c.

The Lorenz curve in Figure 6 is called the "curve of absolute inequality." It is an abstraction and represents the theoretical curve of "maximum sag." It is not very likely that any real Lorenz curve will look like the one in Figure 6. However, the curve for Bolivia, drawn in Exercise 5 comes close. (See Table V.)

We have now established the limiting cases for Lorenz curves, and these are shown in summary in Figure 7. Any real Lorenz curve (such as the dashed curve in Figure 7) lies in the triangular region between the curve of absolute inequality. The closer the Lorenz curve is to the curve of absolute equality, the closer to equality is the distribution of resources it represents; the closer it is to the curve of absolute inequality, the closer the distribution is to inequality.

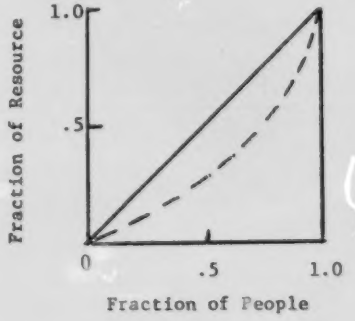


Figure 7.

The Lorenz curve provides us with a way to visualize and compare the distributions of resources. But a Lorenz curve alone gives only a *qualitative* measure of how equally a resource is distributed, and we really need some *quantitative* measures. In the two following sections of this module, we set up some quantitative measures based on the Lorenz curves we have been examining in this section.

1.4 Exercises

Exercise 6.

Table VII shows how population was divided among the 25 largest countries in 1963 and 1968.* Was there much change in the population distribution between 1963 and 1968? How would the Lorenz curves for 1963 and 1968 compare? Graph each using the same set of coordinates.

Table VII

Fraction of Countries	Fraction of Population	
	1963 (x)	1968 (●)
0.0	0.0	0.0
.2	.046	.048
.4	.105	.107
.6	.193	.194
.8	.345	.347
1.0	1.000	1.000

Question for Discussion:

How do you reconcile the data in Table VII (and in the resulting graph) with the fact that world population is growing rapidly?

Exercise 7.

Table VIII shows how total energy produced and total energy consumed was divided among the 25 largest producers and users of energy in 1963.* Draw the corresponding Lorenz curves on the same set of coordinates. Was production of energy more or less equally divided than was use of energy?

*Data from Comparative Analysis of Political Environments, as quoted by Harf, Ledford, and Thompson, op. cit.

*Data from Comparative Analysis of Political Environments, as quoted by Harf, Ledford, and Thompson, op. cit.

Table VIII

Fraction of Countries	Fraction of Energy	
	(x) Used	Produced (o)
0.0	0.00	0.00
.2	.027	.032
.4	.065	.084
.6	.135	.170
.8	.266	.3031
1.0	1.000	1.000

Exercise 8.

Tables I and III are for the distribution of income in the United States in 1955 and 1971, respectively. Compare the data in these two tables (also analyse the corresponding Lorenz curves). Can you say whether income was distributed more or less equally in 1971 than it was in 1955? Discuss.

Exercise 9.

Letting p stand for the fraction of the population and r for the fraction of the resource, consider the family of Lorenz curves given by the equation

$$r = p^n, \quad n = 1, 2, 3, 4, \ldots$$

a) To what situation does the case n = 1 correspond? (See Exercise 4.)

b) In what way is the number n a measure of the equality of distribution? To answer this question, plot $r = p^n$ $(0 \leq p \leq 1)$ for $n = 2, 4, 8, 10$. (Although n is a quantitative measure of equality, it is not generally useful because Lorenz curves are certainly not often of the form $r = p^n$.)

c) To what situation does the Lorenz curve $r = p^n$ correspond in the limiting case $n \to \infty$? To answer the question try drawing the curve $r = p^{100}$ for $0 \leq p \leq 1$.

Exercise 10 (Advanced).

Can a Lorenz curve have the shape shown in Figure 8? Discuss.

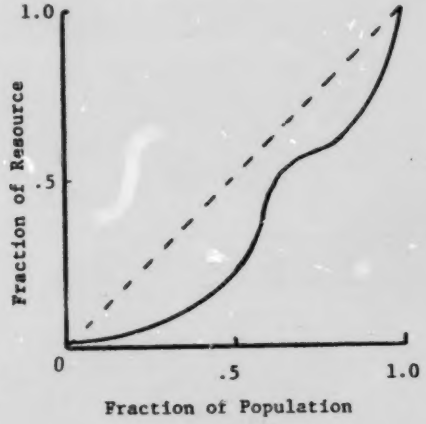


Figure 8.

1.5 Model Exam

The accompanying table shows how gross national product (GNP) is distributed among the countries of the world:

Table IX

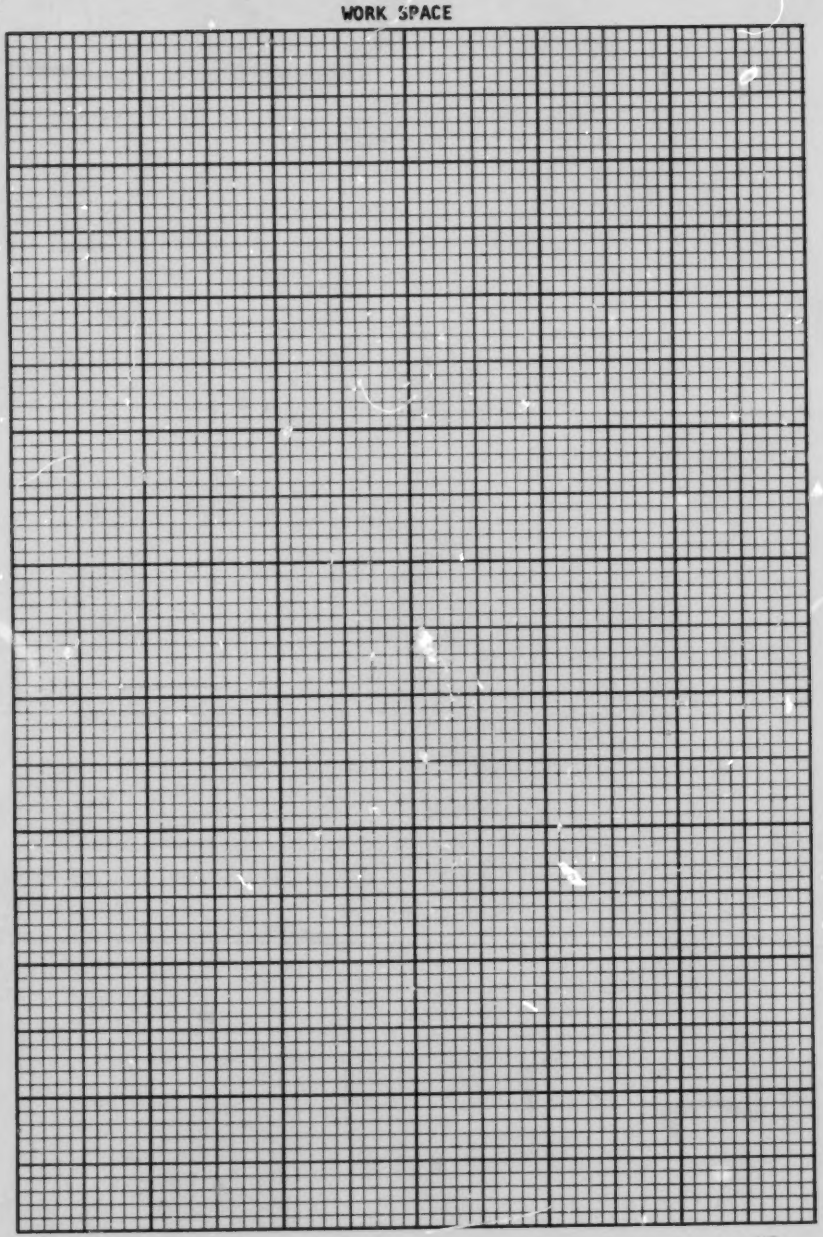
Fraction of Countries	Fraction of GNP
0.0	0.0
.1	.001
.2	.002
.3	.005
.4	.010
.5	.018
.6	.028
.7	.058
.8	.11
.9	.21
1.0	1.00

1. In the space provided sketch a graph of the Lorenz curve for the data in this table. Label the axes carefully.

2. On the same graph used in question 1, draw the Lorenz curve that corresponds to a *completely unequal* distribution of GNP, and the Lorenz curve that corresponds to a *completely equal* distribution of GNP. Label each.

3. On the basis of the curves drawn in question 1 and 2 would you say that GNP is divided equally or unequally among the countries of the world? Write a brief paragraph in support of your answer.

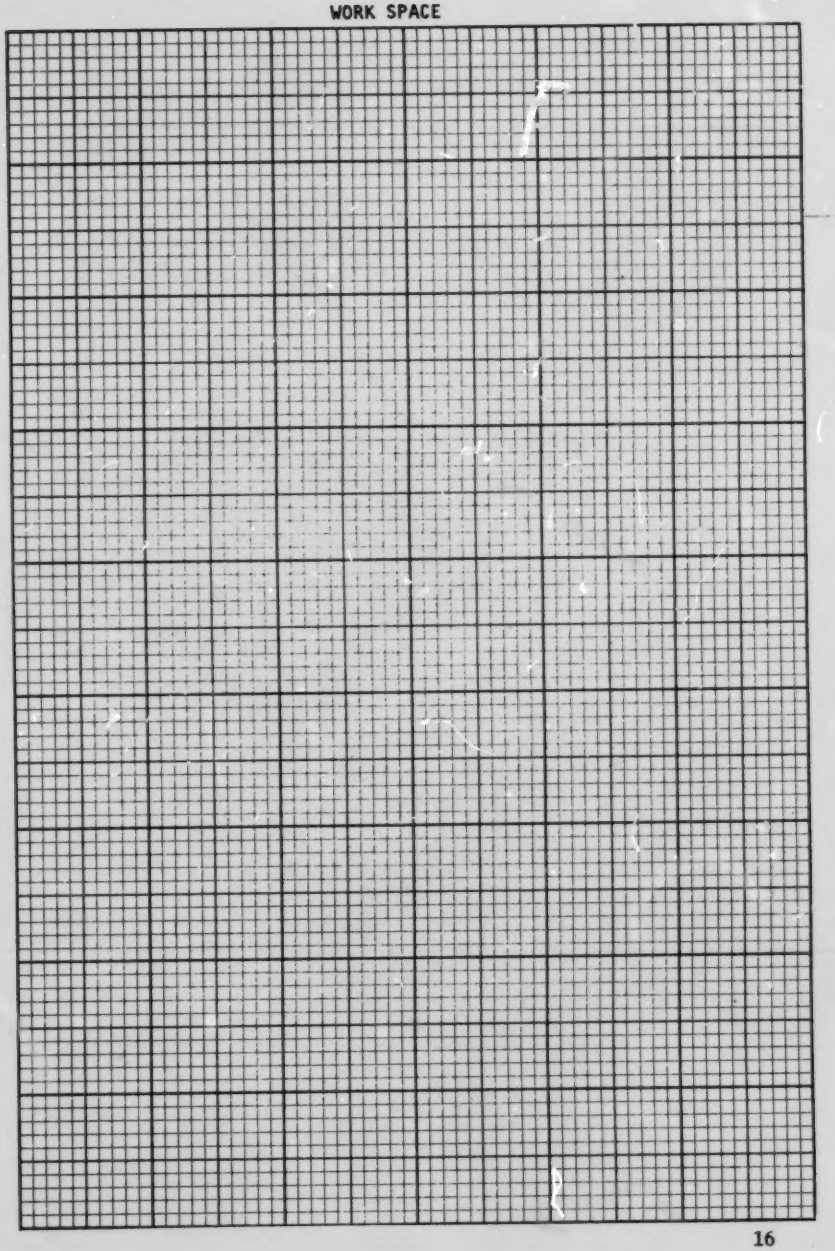
WORK SPACE



15

20

WORK SPACE



16

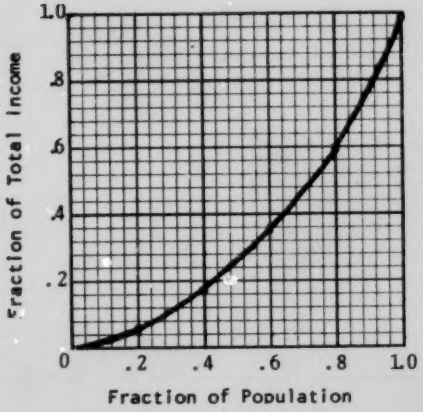
21

1.6 Answers to Exercises

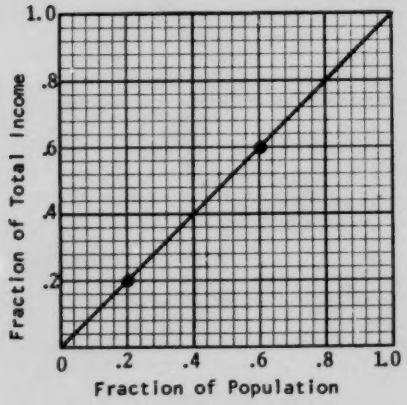
1.

Fraction of People	Fraction of Income
0.1	0.02
.35	.13
.5	.24
.75	.48
.9	.75

2.

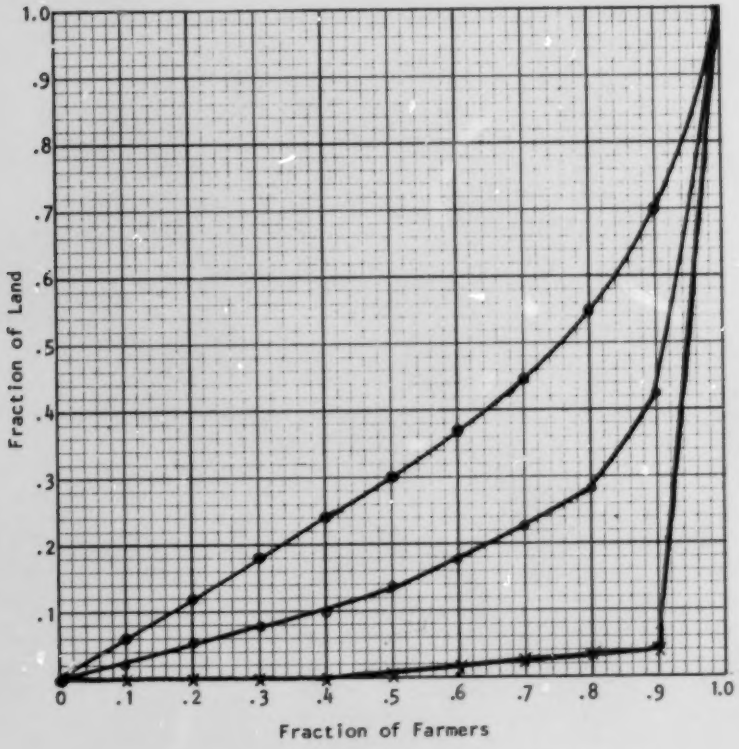


3.



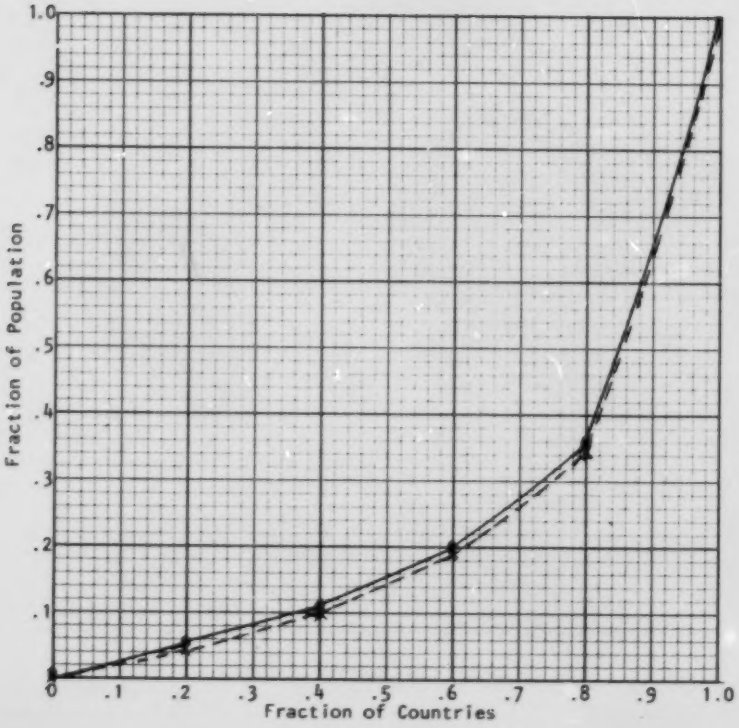
4. r = p

5.



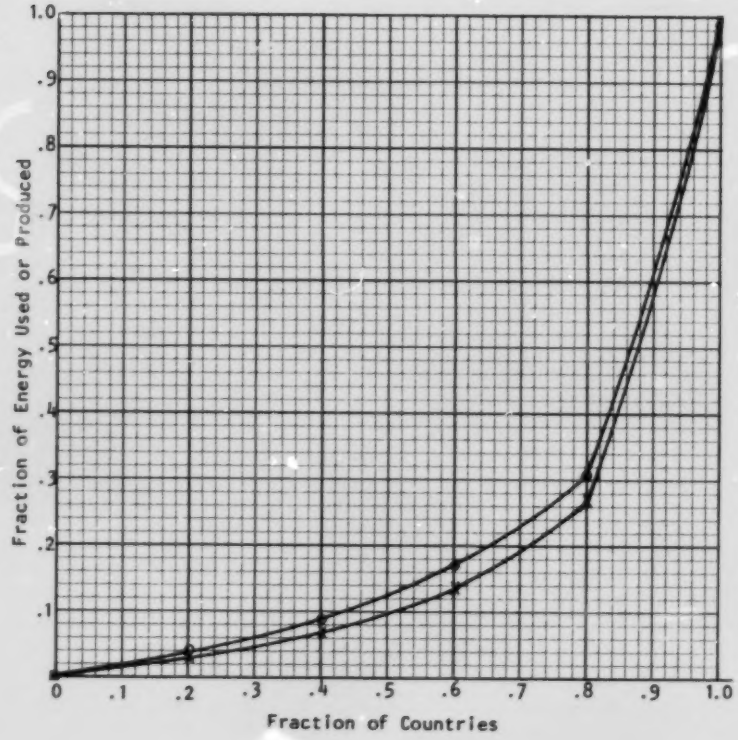
x (Bolivia) • (Denmark) o (United States)

6.



x (1963) • (1968)

7.



8. More equally in 1971.

9. a) Complete equality $r = p^1 = p$.

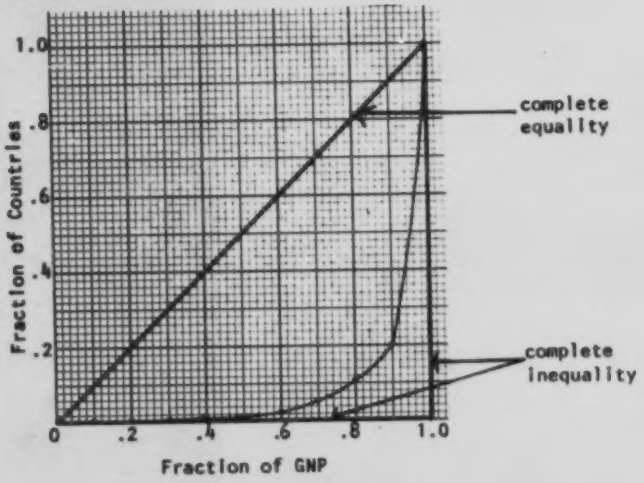
   b) The larger the n, the less equal the distribution. Note the larger the n, the more of a "corner" near 1.

   c) Complete inequality.

10. No, the smallest values are at the start of the curve. We must put the smallest shares at the beginning so the slope is always increasing as we go from left to right.

1.7 Answers to Model Exam

1 and 2.



3. Very unequal. The Lorenz curve is very close to the curve of complete inequality.

2. THE EQUAL SHARE COEFFICIENT

2.1 Introduction

Section 1 of this module discussed Lorenz curves and how they can help us to visualize the way in which resources are distributed. But a Lorenz curve by itself can give us only a qualitative measure of equality. Several different methods have been invented which provide quantitative measures of equality. In this Section, we study one of these; it is called the equal share coefficient (ESC) and involves the mathematical concepts of slope and derivative.

2.2 The Slope of a Lorenz Curve

In order to define the ESC, we need to learn the significance of the slope of a Lorenz curve. To do this, we go back to the 1955 distribution of income studied in Section 1. The Lorenz curve for this distribution is repeated as Figure 1 of this Section.

The point marked A in Figure 1 corresponds to the lowest .3 of the population and .10 of the total income. The point B corresponds to the lowest .4 of the population and .16 of the income. Thus, the .1 of the population between the lowest .3 and the lowest .4 receives .16 - .10 = .06 of the total income. That is, this .1 of the population, since it gets only .06 of the income, receives *less* than an equal share. (An equal share for .1 of the population would, of course, be .1 of the income.) Notice that the slope of the line through the points A and B is .6; that is

$$\begin{bmatrix}\text{slope} \\ \text{of AB}\end{bmatrix} = \frac{.16 - .10}{.4 - .3} = \frac{.06}{.1} = .6.$$

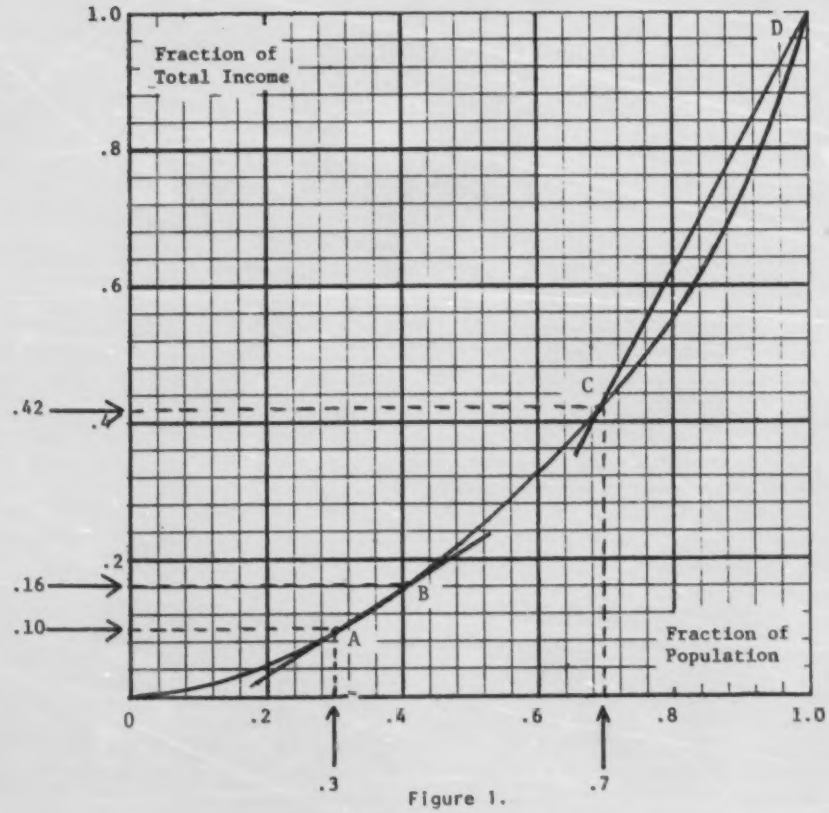


Figure 1.

This slope is less than 1, and as we see, that is a geometric way of saying that the part of the population between .3 and .4 gets less than an equal share of the income.

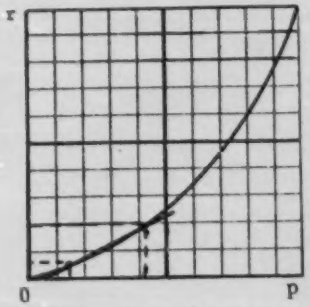
Now look at two other points on the curve, C and D. The point C corresponds to the lowest .7 of the population and .42 of the income, and D corresponds to the whole population (1) and the whole income (1). Thus, this 1 - .7 = .3 of the population earns 1 - .42 = .58 of the income which is *more* than an equal share. (An equal share for .3 of the population would be .3 of the income.) Notice that the slope of

the line CD is approximately 1.93:

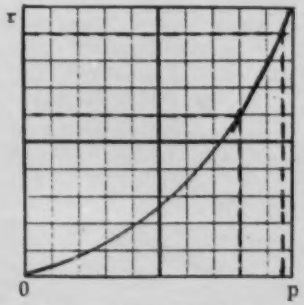
$$\begin{pmatrix}\text{slope}\\ \text{of CD}\end{pmatrix} = \frac{1 - .42}{1 - .7} = \frac{.58}{.3} = 1.93.$$

This slope is greater than one, which is a geometric way of saying that the .3 of the population between .7 and 1.0 gets more than an equal share of the income.

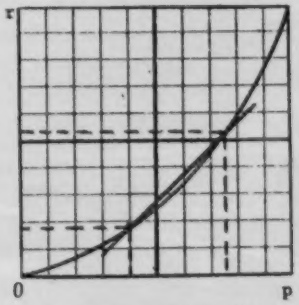
In general, the slope of a line joining two points on a Lorenz curve is less than one if the fraction of the population between these points gets less than an equal share of the income (Figure 2a). The slope is greater than one if the fraction of the population gets more than an equal share (Figure 2b). The slope is equal to one if the fraction of the population receives exactly an equal share (Figure 2c).



a) slope less than 1; less than an equal share.



b) slope greater than 1; more than an equal share.



c) slope equal to 1; an equal share.

Figure 2.

Exercise 1.

Figure 3 is the Lorenz curve for the distribution of land among farmers in the United States (see Section 1, Table V, Problem 5). By measuring slopes from this curve, determine whether each of the following groups of farmers own more or less than an equal share of the land:

a. The .1 between .1 and .2

b. The .2 between .3 and .5

c. The .05 between .6 and .65

d. The .1 between .9 and 1.0

Exercise 2.

Use Figure 3 to find the .2 of the farmers who own an exactly equal share of the land. (Your answer, since it will be determined from a graph, can only be approximate.)

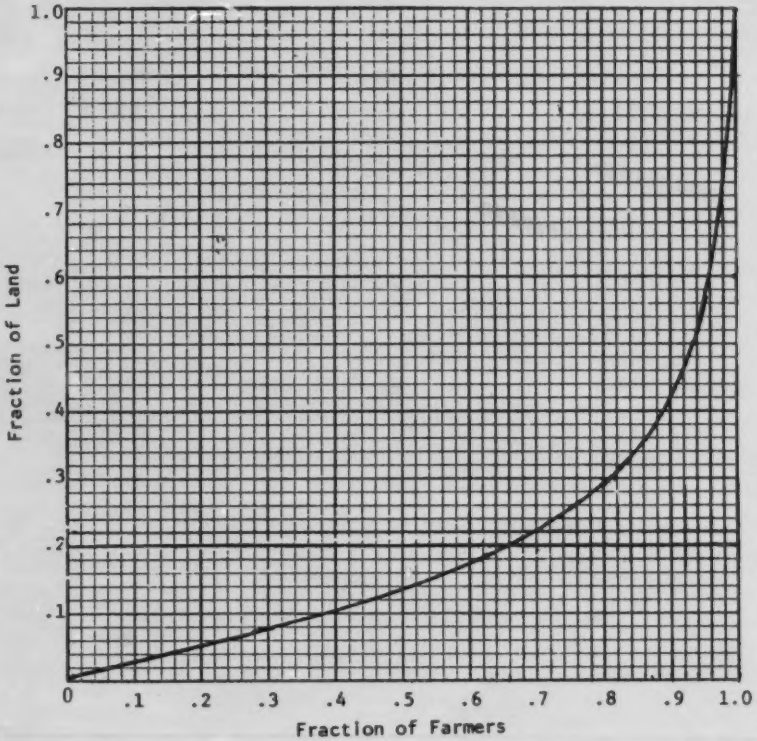


Figure 3.

2.3 The Definition

So far our discussion has dealt only with the *average* slope of Lorenz curves (that is, a slope over some interval between two points). This is not really precise enough a quantity for our purposes for the following reason. Figure 4a shows a hypothetical Lorenz curve for some resource. You should be able to convince yourself that the .2 of the population between .4 and .6 has an exactly equal share (.2) of the resource (draw the line whose slope shows this). The trouble with this statement is that it is only true in an *average* sense: A group on the lower end of this .2 of the population (as for example, the .1 between .4 and .5) gets *less* than an equal share, and a group near the upper end (between .5 and .6 for example) gets *more* than an equal share. The .2 of the population between .4 and .6 gets an equal share only on the average.

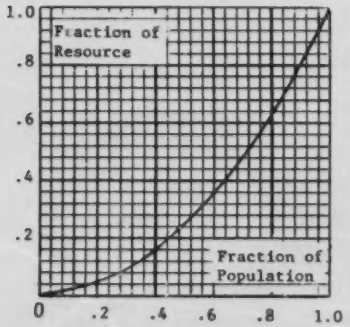


Figure 4a.

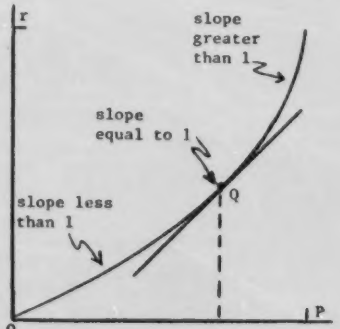


Figure 4b.

A more precise way to measure equality of distribution is provided by the *single point* on the Lorenz curve at which the slope becomes equal to 1. That point Q is indicated in Figure 4b. The slope of the curve everwhere to the left of that point is less than one; everywhere to the right, it is greater than 1. (This is because a Lorenz curve is concave upwards and monotonically increasing.) The value of p at which the slope of the Lorenz curve becomes 1 is called the *equal share coefficient* (ESC) which we denote by ε. The ESC is the fraction of the population that receives less than an equal share of the resource. Note that $0 \leq \varepsilon \leq 1$.

If, for example, $\varepsilon = .4$, that means that .4 of the population owns less than an equal share of the resource, but if $\varepsilon = .75$, then .75 of the population owns less than an equal share of the resource.

If we denote a Lorenz curve by $r = f(p)$, then ε is the (smallest) value of p for which

$$\frac{dr}{dp} = 1,$$

since the derivative is a measure of the slope. Below we show a few idealized cases in which r is a given function of p. In those cases, we calculate ESC's by differentiation. As a rule, however, we have tables or graphs but no formulas and cannot calculate formulas for derivatives. In those cases, we have to work geometrically or arithmetically as we now show.

2.4 A Geometric Approach

First we discuss the geometric way to determine ε. This method is most useful when we have a carefully drawn Lorenz curve. The procedure is illustrated in Figures 5a-e. Place a plastic drafting triangle so that one leg passes through the points (0,0) and (1,1) of the Lorenz curve (Figure 5a). Note that the slope of this leg is 1. Now place a ruler against the base of the triangle (Figure 5b). Hold the triangle firmly as you do this so it cannot move. Then, holding the ruler firmly, begin to slide the triangle to the right (Figure 5c). Note that the slope of the triangle leg remains

equal to one, since, by holding the ruler firmly, you ensure that the triangle does not rotate as it slides. Continue sliding the triangle unitl the leg just touches or grazes the Lorenz curve at a single point (Figure 5d). Mark that point on the curve. Finally, remove the triangle and ruler (Figure 5e) and read the horizontal coordinate of the point. That coordinate is ε, the ESC.

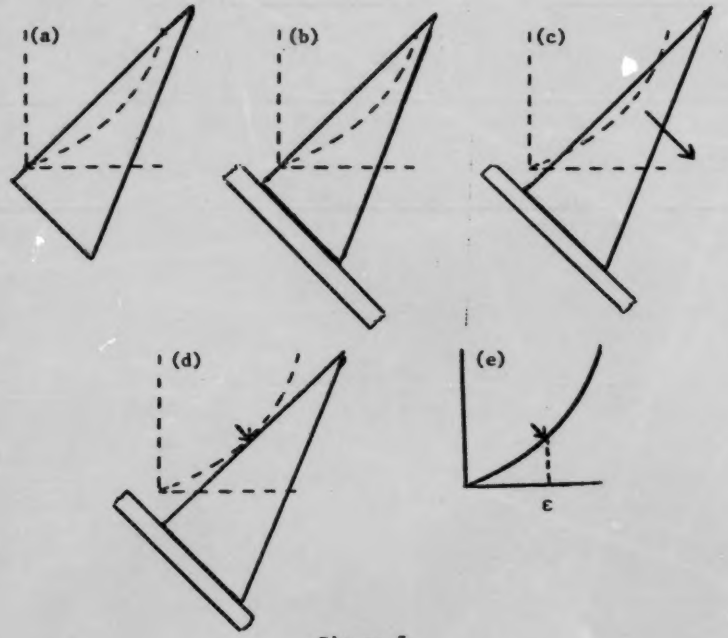


Figure 5.

Exercise 3.

Use the graphical method described above to determine the ESC for each of the following Lorenz curves from Section 1:

a. Figure 1.
b. Answer to Exercise 2.
c. Answer to Exercise 5.
d. Answer to Exercise 6.
e. Answer to Exercise 7.

Exercise 4.

Using part (c) of Exercise 3 above as a guide, state whether, in general, a value of ε greater than .5 indicates an equal or an unequal distribution.

Exercise 5.

Figure 6 is the Lorenz curve for a perfectly equal distribution. Note that the slope of this curve is one for all values of p. What value would you assign to the ESC here? (Remember ε is the fraction of the population that receives less than an equal share of the resource.)

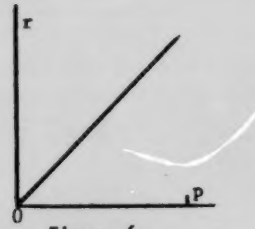


Figure 6.

Exercise 4 touches upon an important issue. In what way is ε a measure of equality? Put in different terms, of two distributions does the one with the larger ESC represent a more or less equal distribution? To do this, we consider two extreme cases. In the most equal of all possible distributions, represented by the curve of absolute equality (Figure 6), everyone gets an equal share, so no one has less than an equal share. The fraction of the population with less than an equal share is therefore zero and so, for perfect equality, $\varepsilon = 0$. On the other hand, if a resource is distributed in a very unequal way (as in Figure 7) then almost everyone gets less than an equal share and ε comes very close to 1.

Figure 7.

These examples suggest that an ESC close to zero implies a large degree of equality in a distribution, while an ESC close to 1 implies a large degree of inequality.

Exercise 6.

Check your results in Exercise 3 to determine if they bear out the assertion made above that a value of ε near zero implies a high degree of equality while one close to zero implies a high degree of inequality.

Exercise 7.

Using your results from Exercise 6, correlate in a qualitative way the value of ε for a Lorenz curve, with the amount that the curve sags below the curve of absolute equality.

2.5 A Numerical Approach

Next, we turn to the numerical method for calculating an ESC. We know that if r is given as a function of p, $r = f(p)$, the ESC is the (smallest) value of p for which

$$\frac{dr}{dp} = 1.$$

An approximate (though often quite accurate) way to calculate such a derivative is illustrated in Figure 8. The derivative at some point $p = p_0$ is approximated by the average slope between $p_0 - h$ and $p_0 + h$. If h is small enough, this should be a good approximation

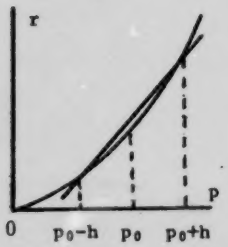


Figure 8.

because by the definition of the derivative

$$r'(p_0) = \lim_{h \to 0} \frac{r(p_0 + h) - r(p_0 - h)}{2h}.$$

Thus

$$r'(p_0) \approx \frac{r(p_0 + h) - r(p_0 - h)}{2h}$$

as long as h is small.

We'll show how this works in the case of the 1955 income distribution quoted in Section 1, Table I. Note that, for example,

$$r(.4) = .16$$

and

$$r(.6) = .33.$$

Thus, according to our approximation (with $h = .1$)

$$r'(.5) = \frac{r(.6) - r(.4)}{2(.1)} = \frac{.33 - .16}{.2} = .85.$$

The numbers from Table I, Section 1 are repeated in Table I of this section and values for r' are also included, calculated as above.

Exercise 8.

Use the formula $r'(p_0) = \frac{r(p_0 + h) - r(p_0 - h)}{2h}$ to verify the numbers in Column 3 of Table I.

We are interested in the value of p for which $r' = 1$. The value 1 does not, however, appear in Table I, so let's make a graph of r' vs. p as in Figure 9, where we have plotted the five points given in Table I and then drawn a smooth curve through them. The graph shows that $r'(p) = 1$ at $p \approx .61$ which is thus

our estimate of the ESC. How does this compare with the value you obtained by the graphical method in Problem 3?

Table I

p	r	r'
0	0	
.1		.25
.2	.05	
.3		.55
.4	.16	
.5		.85
.6	.33	
.7		1.1
.8	.55	
.9		2.25
1.0	1.0	

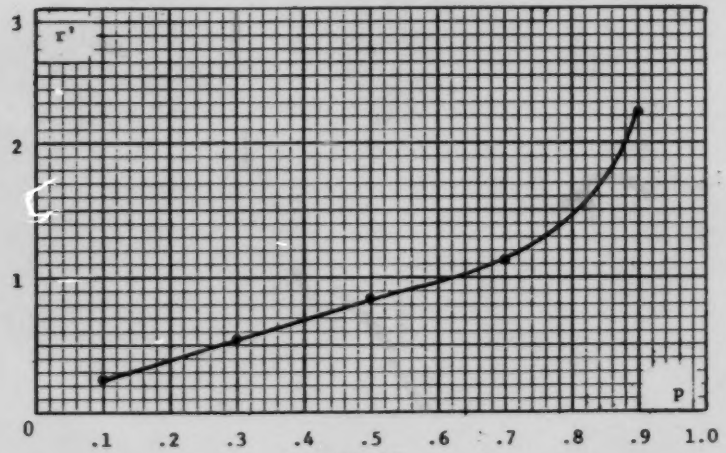


Figure 9.

Exercise 9.

Using the method just described, calculate and plot r' vs. p for the distributions given in Tables III, V, and VI of Section 1, and determine a value for the ESC in each case. Compare your results with those obtained by the sliding triangle method. Which method, the present one, or the sliding triangle, do you believe gives the more accurate results? Discuss.

2.6 Linear Interpolation

There is another procedure for finding ε that does not involve reading numbers from a graph. It is called the method of linear interpolation. In Figure 10, we plot the two points from Table I which bracket the value r' = 1 (that is, p = .5, r' = .85, and p = .7, r' = 1.1) and then connect these two points with a straight line.

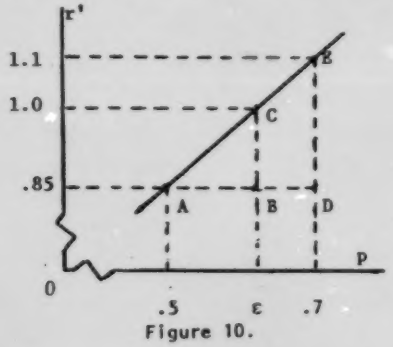


Figure 10.

This straight line is an approximation to the curve r' vs. p in the interval between p = .5 and p = .7. The underlying idea (or pious hope) is that any well-mannered curve will be approximately straight over a short interval. In any case, ε is the value of p at which r' = 1. To find ε we use the fact that triangles ABC and ADE in Figure 10 are similar:

$$\frac{AB}{AD} = \frac{BC}{DE} .$$

However,

$$AB = \varepsilon - .5$$
$$AD = .7 - .5 = .2$$
$$BC = 1.0 - .85 = .15$$
$$DE = 1.1 - .85 = .25$$

so

$$\frac{\varepsilon - .5}{.2} = \frac{.15}{.25}$$

or

$$\varepsilon = .5 + .2\left(\frac{.15}{.25}\right) = .62$$

which is close to the value we read from Figure 9.

2.7 Exercises

Exercise 10.

Recalculate the ESC's found in Problem 9 by the method of linear interpolation. How do these values compare with those calculated in Problem 8? Which set do you think is the more accurate? Discuss.

Exercise 11.

a) Let $r = p^n$ $(n > 1)$ be the equation of a Lorenz curve. Find an expression for the ESC in terms of n.

b) Use the result of (a) to calculate ε for $n = 2, 4, 8, 16$. Are these results reasonable (see Section 1, Problem 9)?

c) Calculate $\lim_{n\to\infty} \varepsilon$ using for ε the formula you derived in (a).

Exercise 12. (Advanced).

If a Lorenz curve is given by $r = f(p)$, the ESC is the value of p for which

$$\frac{dr}{dp} = f'(p) = 1.$$

But how can we be sure that $f'(p)$ will equal 1 at some p between 0 and 1? Making the necessary assumptions about the continuity and differentiability of $f(p)$, use the mean value theorem of differential calculus to prove that $f'(p) = 1$ for at least one value of p between 0 and 1.

Exercise 13.

Let a Lorenz curve be given by

$$r = \frac{1}{2}p^{\frac{3}{2}}(1 + p).$$

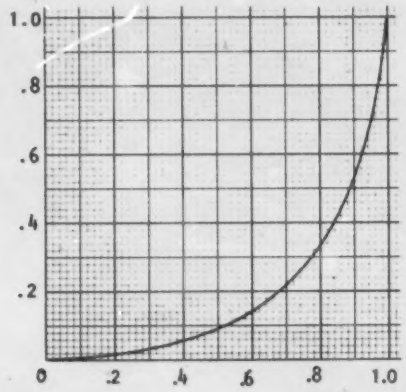
Show that the ESC ε is a solution of

$$\frac{3}{4}\varepsilon^{\frac{1}{2}} + \frac{5}{4}\varepsilon^{\frac{3}{2}} = 1.$$

2.8 Model Exam

1. Define the equal share coefficient, and discuss how it measures the equality of a resource distribution.

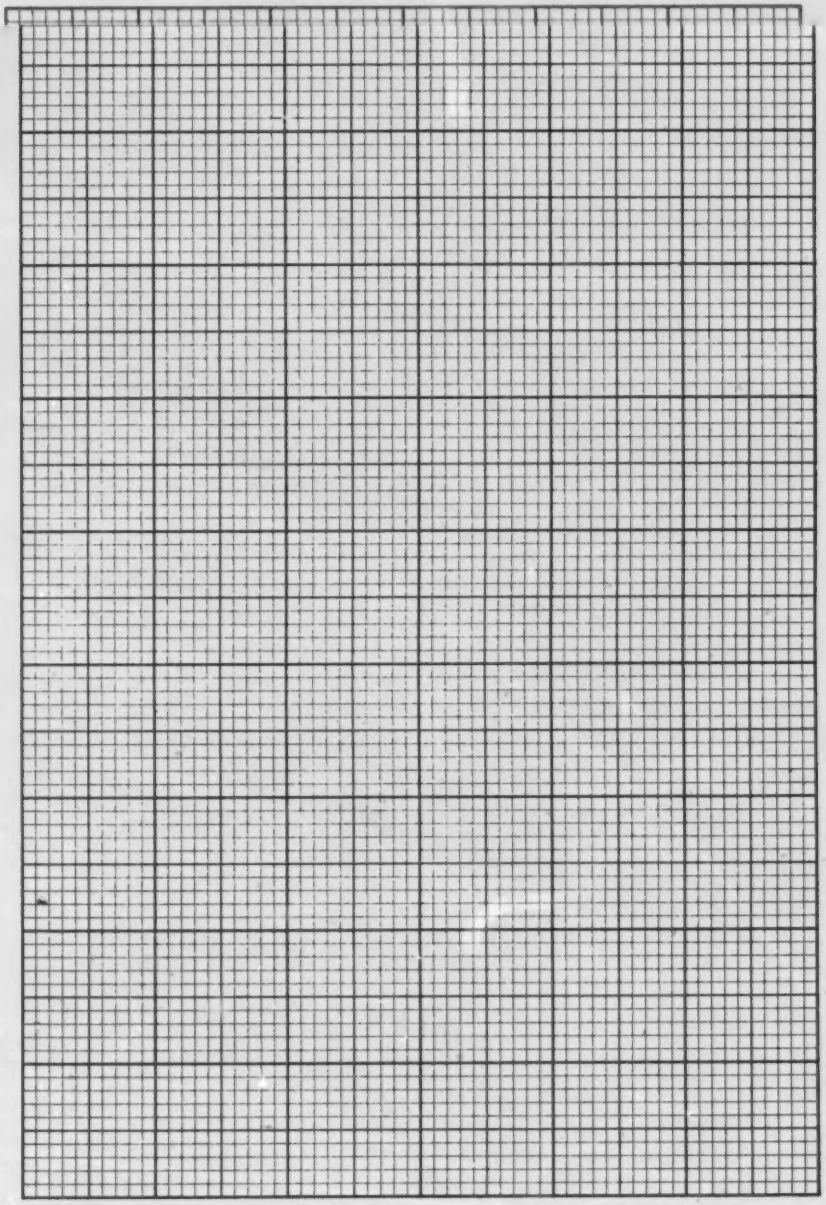
2. In the Lorenz curve shown below, does the fraction of the population p between p = .3 and p = .4 own more than an equal share, less than an equal share, or an equal share of the resource r? Discuss how you arrive at your answer.



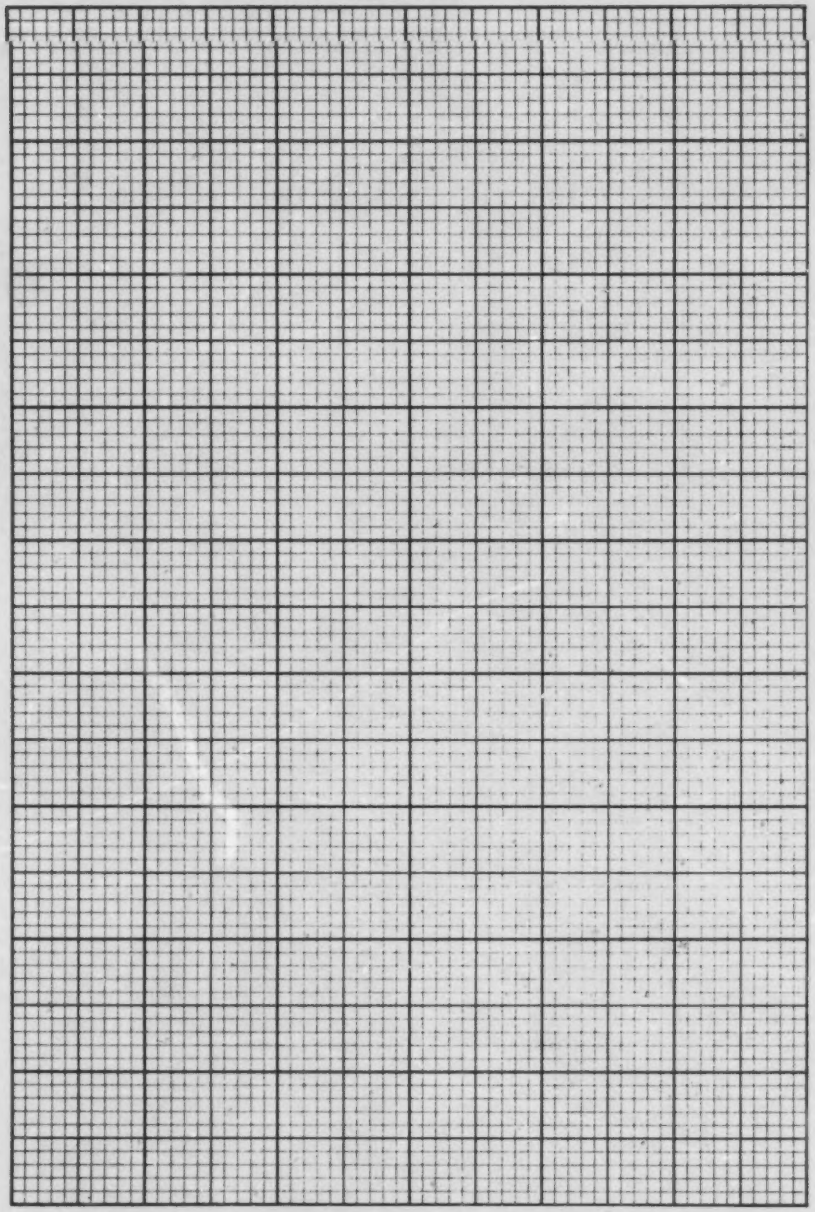
3. Estimate the equal share coefficient for the Lorenz curve shown above by graphical means. Show your work.

4. A Lorenz curve is given by $r = \frac{1}{2}(p+p^2)$. Find the equal share coefficient. On the basis of this calculation and your answer to Question 3, state which distribution is more nearly equal, the one shown in the graph above, or the one given by this equation. Explain.

WORK SPACE



WORK SPACE



2.9 Answers to Exercises

1. a) less than
   b) less than
   c) less than
   d) more than

2. The .2 between .7 and .9.

3. a) .64
   b) .6
   c) .8, .82, .9
   d) .8, .8
   e) .8, .8

4. Values larger than .5 indicate an unequal distribution in favor of the few near the top.

5. $\varepsilon = 0$

6. All answers in Exercise 3 are greater than .5. This, as was asserted, indicates a high degree of inequality in all cases.

7. $0 \leq \varepsilon \leq 1$. $\varepsilon = 0$ = absolute equality.
   $\varepsilon = 1$ = absolute inequality.
   $\varepsilon = 0$ means no sag
   $\varepsilon = .5$ means a sag half way
   $\varepsilon = 1$ means complete sag

   The larger the ε, the more the curve will sag.

8. $r'(.1) = \frac{r(.2) - r(0)}{2(.1)} = \frac{.05 - 0}{.2} = .25$

   $r'(.3) = \frac{r(.4) - r(.2)}{2(.1)} = \frac{.16 - .05}{.2} = .55$

   $r'(.7) = \frac{r(.8) - r(.6)}{2(.1)} = \frac{.55 - .33}{.2} = 1.1$

   $r'(.9) = \frac{r(1.0) - r(.8)}{2(.1)} = \frac{1.0 - 5.5}{.2} = 2.25$

9.

Table III'

p	r	r'
0	0	
.1		.3
.2	.06	
.3		.6
.4	.18	
.5		.9
.6	.36	
.7		1.2
.8	.60	
.9		2.0
1.0	1.00	

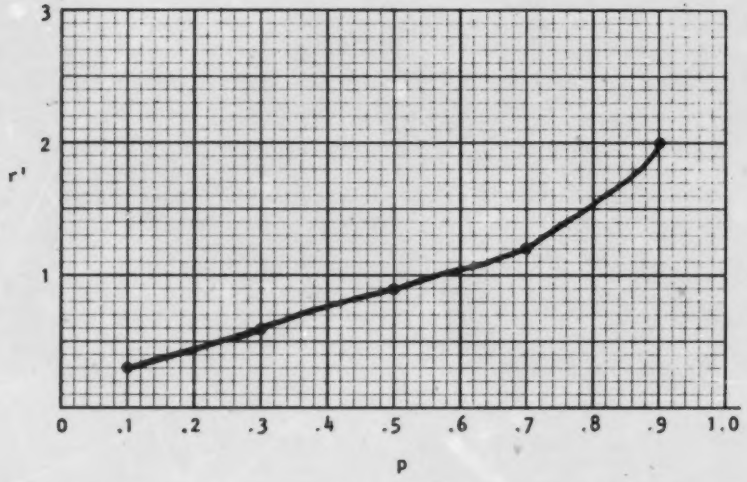
$$r'(.1) = \frac{r(.2) - r(0)}{2(.1)} = \frac{.06}{.2} = .3$$

$$r(.3) = \frac{r(.4) - r(.2)}{2(.1)} = \frac{.18 - .06}{.2} = .6$$

$$r(.5) = \frac{r(.6) - r(.4)}{2(.1)} = \frac{.36 - .18}{.2} = .9$$

$$r(.7) = \frac{r(.8) - r(.6)}{2(.1)} = \frac{.60 - .36}{.2} = 1.2$$

$$r(.9) = \frac{r(1.0) - r(.8)}{2(.1)} = \frac{1.0 - .6}{.2} = 2.0$$

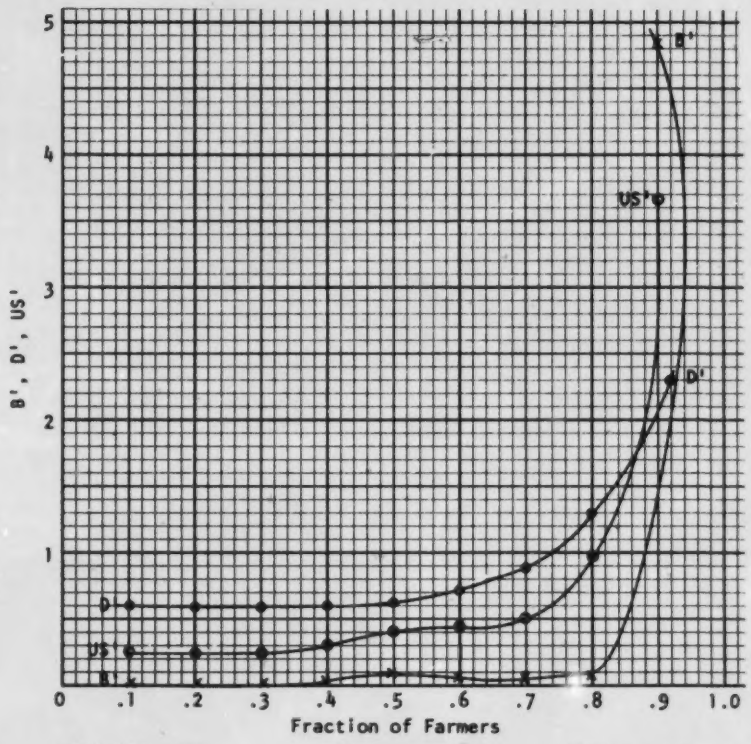


ESC = .6
Triangle method also gives .6.

9. (cont.)

Table V'

Fraction of Farmers	Fraction of Land (1964)					
	B(x)	B'	D(•)	D'	US(o)	US'
0.0	0		0		0.0	
.1	0	0	.06	.60	.025	.25
.2	0	0	.12	.60	.05	.25
.3	0	0	.18	.60	.075	.25
.4	0	.02	.24	.60	.10	.275
.5	.01	.08	.30	.60	.13	.40
.6	.016	.06	.36	.75	.18	.45
.7	.022	.07	.45	.90	.22	.5
.8	.03	.09	.54	1.25	.28	1.0
.9	.04	4.85	.70	2.30	.42	3.60
1.0	1.00		1.00		1.00	

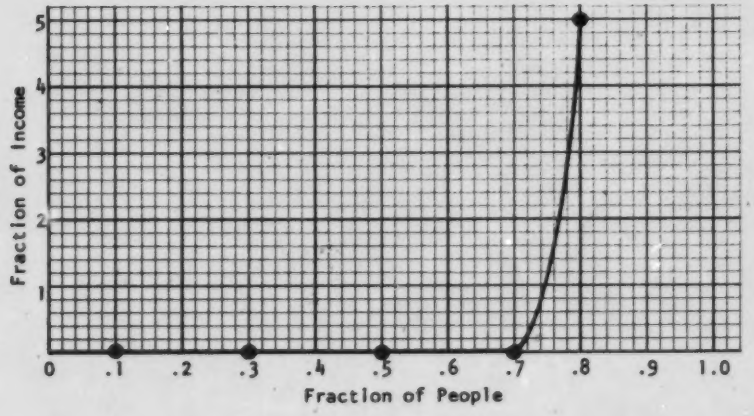


ESC(x) = .812 ESC(•) = .715 ESC(o) = .8

9. (cont.)

Table VI'

Fraction of People	Fraction of Income	r'
0.0	0	
.1		0
.2	0	
.3		0
.4	0	
.5		0
.6	0	
.7		0
.8	0	
.9		5
1.0		



ESC ≈ .84

Both the graphical and the sliding triangle methods give only approximate values.

47

10. For Table III: ESC ≈ .5667

For Table V: ESC(x) ≈ .8191

ESC(o) ≈ .7286

ESC(.) ≈ .8

For Table VI: ESC ≈ .74

11. a) $np^{n-1} = 1.$

b) For $n = 2$, $p = .5$; for $n = 4$, $p = \sqrt[3]{\frac{1}{4}} \approx .63$.

For $n = 8$, $p = \sqrt[7]{\frac{1}{8}} \approx .77$; for $n = 16$, $p = \sqrt[15]{\frac{1}{16}} \approx .83$.

c) $\lim_{n \to \infty} \varepsilon = 1$

12. The mean value theorem says: If $f(x)$ is continuous for $a \leq x \leq b$ and $f'(x)$ exists for $a < x < b$, then there exists at least one value c, such that $a < c < b$, for which

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

For this problem $f(p)$ is continuous on $1 \leq p \leq 0$ and $f'(p)$ exists on $0 < p < 1$. Therefore, by the mean value theorum, we can write

$$f'(p) = \frac{f(1) - f(0)}{1 - 0} = \frac{1 - 0}{1 - 0} = 1$$

for at least one value of p on $0 < p < 1$.

13. a) $r' = \frac{1}{2} p^{\frac{3}{2}} + (1 + p)\frac{3}{4} p^{\frac{1}{2}}$

$$\therefore \frac{1}{2} \varepsilon^{\frac{3}{2}} + (1 + p)\frac{3}{4} \varepsilon^{\frac{1}{2}} = 1.$$

and $$\frac{5}{4} \varepsilon^{\frac{3}{2}} + \frac{3}{4} \varepsilon^{\frac{1}{2}} = 1.$$

b) $\varepsilon \approx .5149$

48

2.10 Answers to Model Exam

1. Given that the Lorenz curve is defined by $r = f(p)$, the value of p at which the Lorenz curve has slope 1 is called the equal share coefficient ε or ESC. The value of ε lies in the interval from 0 to 1, inclusive. A value of ε close to 0 indicates a very nearly equal sharing of resources. A value close to 1 indicates a very unequal sharing.

2. Less than an equal share. The slope is less than $\frac{1}{4}$.

3. Approximately .78.

4. To find ε, find r' and set it equal to 1.

$$r' = \frac{1}{2} + p \quad \therefore \varepsilon = \frac{1}{2} = .5.$$

This distribution is more nearly equal than the one shown on the graph.

3. THE GINI INDEX

3.1 Introduction

Section 2 of this module developed a quantitative measure of the equality of distributions, the equal share coefficient. This quantity is found by locating the point on the Lorenz curve at which the slope is 1. Calculating the ESC is therefore a problem in differential calculus. In this section, we introduce another measure of equality, one which is called the Gini index. To determine the Gini index of a distribution, it is necessary to calculate the area under a Lorenz curve, and so we will be dealing here with a problem in integral calculus.

3.2 Area Under the Lorenz Curve

In Section 1 we learned that in the distribution of any resource, there are two idealized cases, that in which the resource is divided equally among the population (absolute equality), and that in which it is divided completely unequally (absolute inequality). The Lorenz curve for absolute equality is shown in Figure 1a and the Lorenz curve for absolute inequality is shown in Figure 1b.

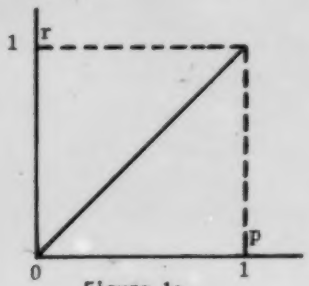


Figure 1a.

Figure 1b.

Lorenz curves for distributions generally lie somewhere between the two extremes pictured in Figures 1a and b. Figures 2a-d show four Lorenz curves illustrating distributions which move further and further from complete

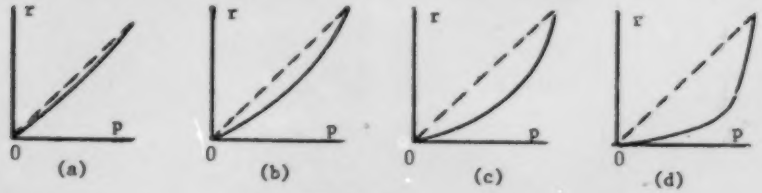


Figure 2.

equality. The area between the curve of absolute equality (the dashed line in Figures 2a-d) and the Lorenz curve grows larger as we move further from absolute equality. This area (cross hatched in Figure 3) is called the *area of inequality*, and we observe that the larger the area of inequality, the less equal the distribution of the resource.

Figure 3.

How big is the area of inequality for the case of absolute equaltiy (Figure 1a)? How big is the area of inequality for the case of absolute inequality (Figure 1b)?

When you determined the area of inequality for the two extreme cases, you should have found that that area is zero for absolute equality and .5 for absolute inequality, as indicated in Figures 4a and 4b. Thus the area

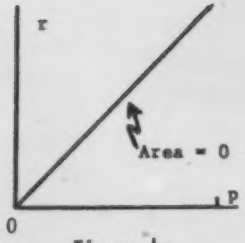


Figure 4a.

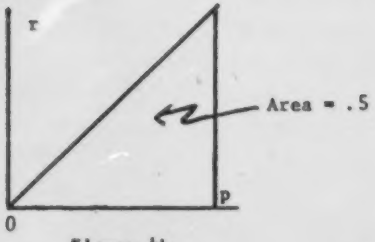


Figure 4b.

of inequality for any real Lorenz curve must lie between 0 and .5.

3.3 The Definition

Although the area of inequality is a measure of the extent to which a resource is divided equally, another quantity closely related to the area of inequality is more often used. This quantity, called the Gini index, is the ratio of the area of inequality to the largest possible area of inequality. In other words,

$$\begin{pmatrix}\text{Gini}\\ \text{Index}\end{pmatrix} = \frac{\begin{pmatrix}\text{Area of}\\ \text{Inequality}\end{pmatrix}}{1/2} = 2 \times \begin{pmatrix}\text{Area of}\\ \text{Inequality}\end{pmatrix}.$$

If we let g stand for the Gini index, and A_I for the area of inequality, then $g = 2A_I$.

Exercise 1.

a) What is the value of g for the case of absolute equality (Figure 1a)?

b) What is the value of g for the case of absolute inequality (Figure 1b)?

Exercise 2.

The Gini indexes for five different distributions of a resource are as follows:

a) .73 c) .41 e) .33

b) .01 d) .86

Arrange these distributions in order from the most to the least equal.

As Exercise 1 shows, a Gini index can be any number from 0 to 1. A value of g close to 0 indicates a distribution close to absolute equality and a value close to 1 indicates a distribution close to absolute inequality.

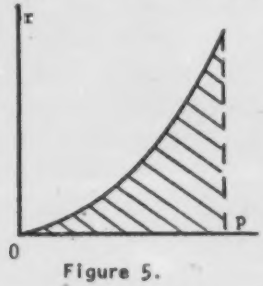


Figure 5.

Exercise 3.

Let A_L stand for the area under the Lorenz curve (Figure 5). Write a formula for the Gini index g in terms of A_L.

Figure 6.

Your result in Problem 3 should be $g = 1 - 2A_L$. Since this expression will be used in what follows to calculate g, it is worth showing how it is derived. Figure 6 is a "diagram equation" which shows that the area under the curve of absolute equality (which is 1/2) equals the area of inequality (A_I) plus the area under the Lorenz curve (A_L). That is,

$$\frac{1}{2} = A_I + A_L.$$

But, as we know, $g = 2A_I$ so $A_I = g/2$, and

$$\frac{1}{2} = \frac{1}{2}g + A_L$$

or

$$\frac{1}{2}g = \frac{1}{2} - A_L.$$

From this we get our formula

$$g = 1 - 2A_L.$$

3.4 Using the Definite Integral

This expression is useful because it reduces the calculation of the Gini index, g, to the calculation of the area under a curve (to be precise, the area under the Lorenz curve), and finding areas under curves is the most famous application of integral calculus. If, as usual, we let r represent the fraction of the resource and p the fraction of the population, then as we see in Figure 7,

$$A_L = \int_0^1 r\,dp$$

So that

$$g = 1 - 2\int_0^1 r\,dp$$

If r is given as a function of p, we can use this expression to find g.

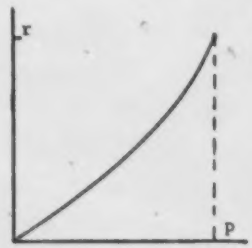


Figure 7.

Exercise 4.

Calculate the Gini index g in each of the following cases:

a. $r = p^2$

b. $r = p^3$

c. $r = p^4$

Which of these represents the most nearly equal distribution? The least?

3.5 The Rectangle Method

Now we calculate g for some actual Lorenz curves. Table 1 shows the distribution of income in the United States in 1955 (this is also Table 1 of Section 1). What is the Gini index for this distribution? We know that we can calculate g once we know A_L, the area under the Lorenz curve. In this situation (and in most situations) we have no formula which gives r in terms of p so we can't calculate $\int_0^1 r\,dp$ exactly by any of the standard "textbook" methods. We can, however, use numerical and graphical methods to approximate the value of the integral.

Table 1

Fraction of People	Fraction of Income
0.0	0.0
.2	.05
.4	.16
.6	.33
.8	.55
1.0	1.00

To show how this can be done, in Figure 8 we draw the Lorenz curve for the distribution in Table 1. Notice that a horizontal line has been drawn on the graph at r = .36. This line creates two triangle-like regions, A and B, which appear to have the same area. If they do have the same area, then the area under the Lorenz curve is equal to the area of the rectangle under the horizontal line. The dimensions of this rectangle are 1 and .36 so its area is 1 x .36 = .36. This is also (approximately) the area A_L under the Lorenz curve in Figure 8. Thus,

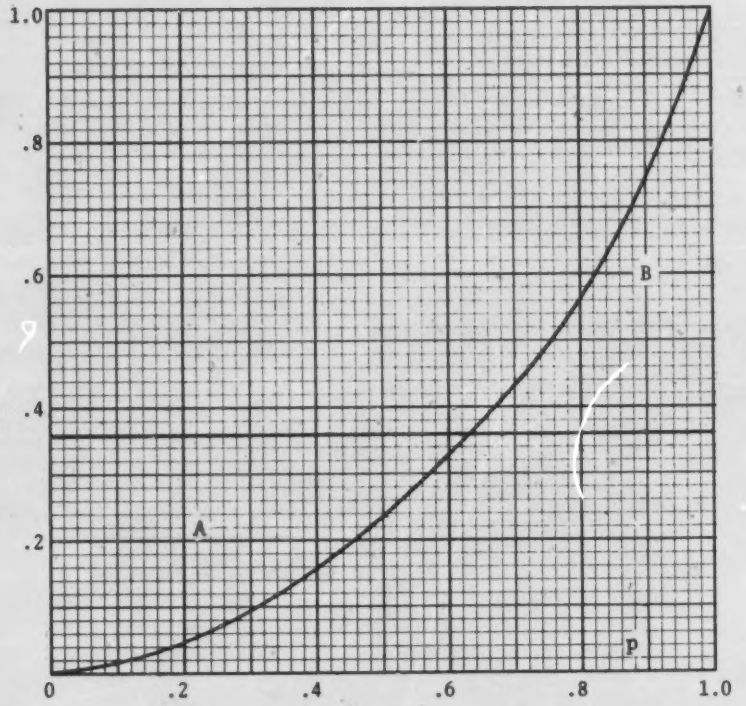


Figure 8.

$g = 1 - 2A_L = 1 - 2(.36) = 1 - .72 = .28.$

In the sense that this number is closer to 0 than to 1, the income distribution in Table 1 is closer to equality than inequality.

It is easy to improve upon this method for calculating g by doing a better job of finding A_L, the area under the Lorenz curve. Figure 9 is a graph of the Lorenz curve also pictured in Figure 8. In Figure 9, however, we have used *two* rectangles, one extending from p = 0 to .5, and the second from p = .5 to p = 1, to estimate the area under the curve. The idea here is that it is easier to

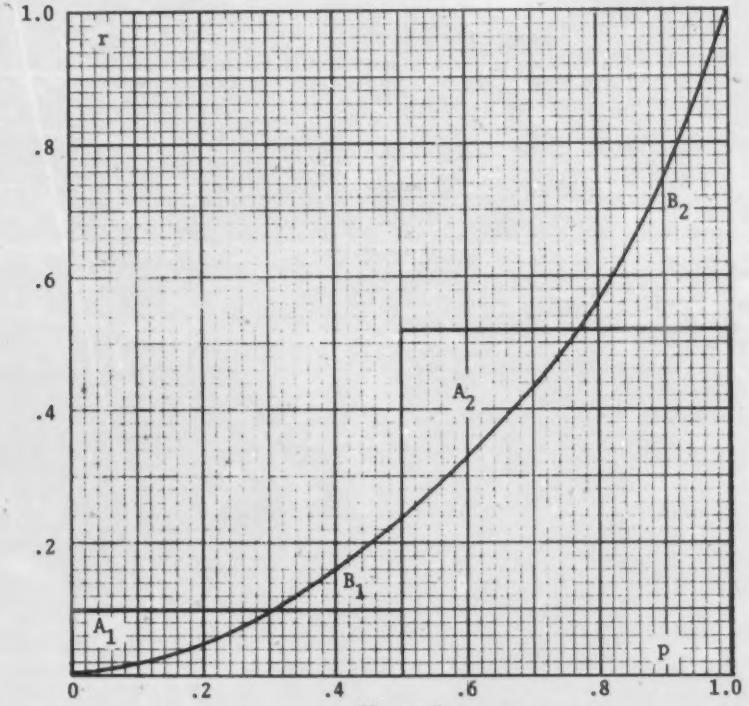


Figure 9.

"eyeball" the height of each of the two narrow rectangles in Figure 9 than it is for the one wide rectangle in Figure 8. The two rectangles in Figure 9 are chosen so that, to the best of our ability to judge by eye, the area of A_1 equals the area of B_1, and the area of A_2 equals the area of B_2.

The height of the left rectangle in Figure 9 is .1 and its width is .5. Thus its area is .05. The area of the right-hand rectangle is .26 = (.52 × .5). Thus, the area of the two rectangles is .05 + .26 = .31. Using our formula for g we find

$$g = 1 - 2(.31) = 1 - .62 = .38,$$

which is substantially larger than the value from the one rectangle approximation.

Exercise 5.

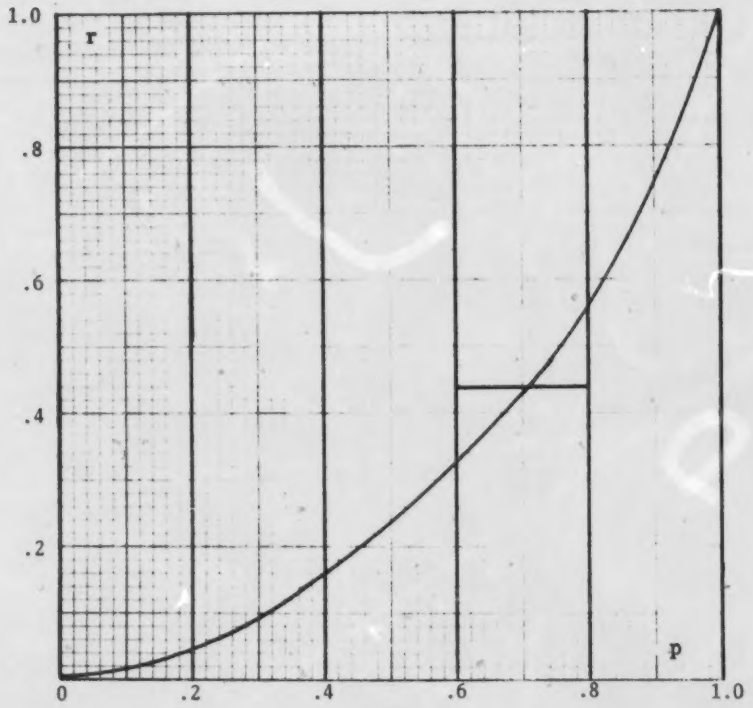


Figure 10.

Figure 10 is the Lorenz curve of Figures 8 and 9. We have divided the graph into five strips each of width .2. We also have drawn a horizontal line in the fourth strip at a height of .44. The area of the rectangle from p = .6 to p = .8 whose height is .44 is .2 x .44 = .088. Estimate the heights of the other four rectangles; compute the areas of each. Then add the areas of the five rectangles to get an approximation to the area A_L under the curve. Use your value A_L to compute g. (Note how this approach is leading us towards the definition of the integral in terms of Riemann sums.)

Exercise 6.

Using five rectangles of equal width, estimate the Gini index for each of the three distributions of land given in Table V, Section 1. Order these three distributions from least to most equal on the basis of their Gini indexes. Does this order agree with the order you made in Exercise 5, Section 1?

Exercise 7.

Use five rectangles of equal width to calculate the Gini index for the distribution of energy production and for the distribution of energy use (data given in Exercise 7, Section 1).

Exercise 8.

a. Draw graphs for each of the three distributions given in Exercise 4 of this Section and make a one-rectangle estimate of the area under each of the curves. Compute the three Gini indexes. How do these values compare with the exact values you found in Exercise 4?

b. Repeat the calculations of (a) using a five-rectangle approximation in each of the three cases. How do these approximate Gini indexes compare with the exact values?

3.6 A Numerical Method

The method of calculating the Gini index we have been discussing involves estimating areas by eye. We now describe another closely related method which is purely numerical and involves no "eyeball" estimation. There is no guarantee that this method provides more accurate answers, but it will permit us to work directly with tabulated values so that we do not need to draw a graph of the Lorenz curve.

To see how this numerical method works, we show, in Figure 11a, a Lorenz curve with the region beneath it divided into strips. Figure 11b shows one of these strips in detail. Our problem is to draw the (dotted) horizontal line which is the top of the rectangle whose

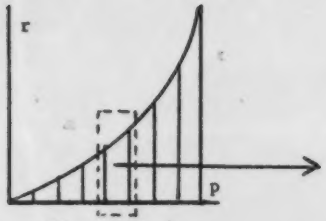


Figure 11a.

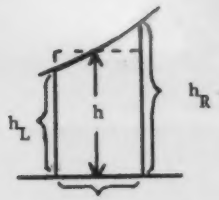


Figure 11b.

area is the same as the area of the strip under the curve. In the graphical method discussed above, this horizontal line was chosen by eye. This method we now propose is to chose this horizontal line so that its height, h, above the horizontal axis is the average of h_L, the height of the curve at the left edge of the strip, and h_R, the height of the right edge. That is,

$$h = \frac{h_L + h_R}{2}.$$

If the width of this strip is w, the area of this rectangle is

$$A = wh = w\left(\frac{h_L + h_R}{2}\right).$$

This area is then an approximation to the area of the strip beneath the curve. If the strip is narrow, the curve's height will be pretty much the same everywhere in the strip (that is h_L and h_R won't differ by much). If that is so, we can choose a rectangle whose height h is *any* value between h_L and h_R and use it to calculate a reasonably good approximation to the area of the strip. The method we are using here is to let h be the average of h_L and h_R. This choice is both sensible and easy to calculate.

To see how this method works, let's use the numbers given in Table I. They are repeated in Table II with some new features added. The column marked "Height (Fraction of Income)" gives the height of the Lorenz curve at each value of p. The column marked "Average Height" is the average of the left and right edge heights. For

Table II

Fraction of People	Height (Fraction of Income)	Average Height	Area of Rectangle
0.0	0.0		
		.025	.005
.2	.05		
		.105	*
.4	.16		
		*	.049
.6	.33		
		.44	.088
.8	.55		
		.775	.155
1.0	1.00		

* See Exercise 9.

example, at .6 the height (fraction of income) is .33 and at .8 it is .55 (see Figure 12). The average of these two numbers is

$$\frac{.33 + .55}{2} = \frac{.88}{2} = .44.$$

This number is entered in the table between the .6 row and the .8 row because it is the average height of the rectangle which extends from .6 to .8.

The area of this rectangle is its height x width. The area is therefore .2 x .44 = .088, and this number appears in the corresponding row of the last column of Table II.

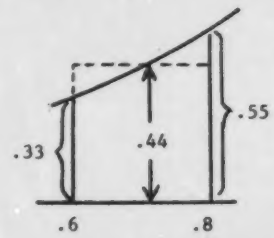


Figure 12.

3.7 Exercises

Exercise 9.

Fill in the missing number in the third column of Table II and the missing number in the fourth column.

The area under the Lorenz curve is approximately the sum of the area of the rectangles in the fourth column of Table II. That is,

$$A_L \approx .005 + (.021) + .049 + .088 + .155 = .318.$$

From this we find

$$g = 1 - 2(.318) = 1 - .636 = .364$$

as the value of the Gini index. Notice that this agrees *very well* with the five strip geometric method we used above.

Exercise 10.

Using the numerical method we have just discussed, calculate the Gini index for each of the three distributions of land given in Table V, Section 1. How do your results compare with those you obtained in Exercise 7? Which set do you think is the more accurate? Discuss.

Exercise 11 (Advanced).

a) Calculate the Gini index g for the Lorenz curve $r = p^n$ where n is a positive number.

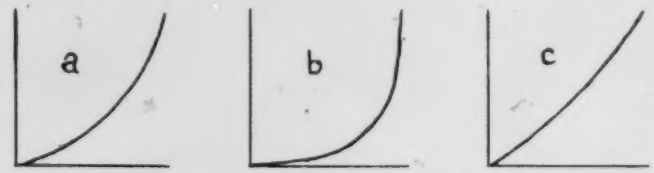
b) Evaluate your result from (a) for the case $n = 1$. Is this the result you would expect? Discuss.

c) Find the limit of g as $n \to \infty$. How can you explain your result?

3.8 Model Exam

1. Define the Gini index and discuss how it measures the equality of a resource distribution.

2. Which of the three distributions whose Lorenz curves are pictured below has the largest Gini index?



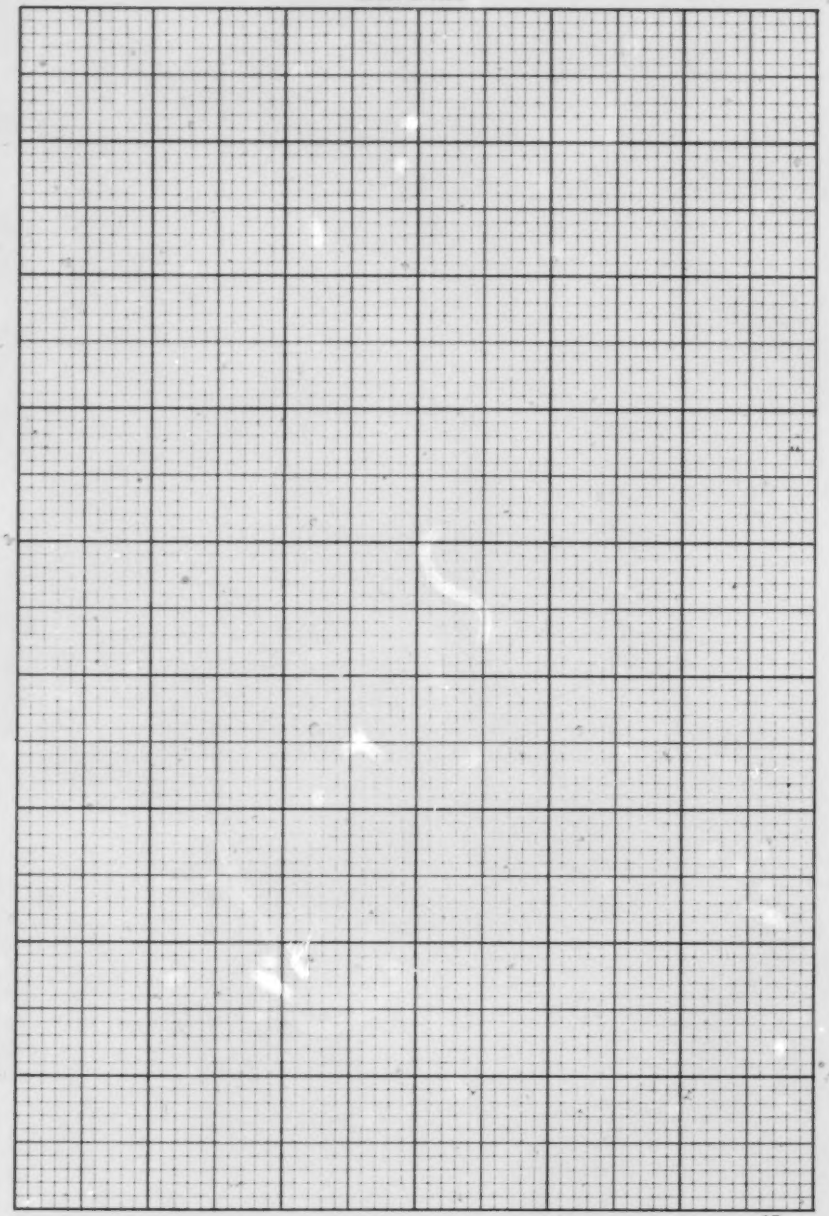
The smallest? Which is the most nearly equal distribution? The least?

3. The accompanying table gives the distribution of a resource. Use these numbers to find a good approximate value of the Gini index.

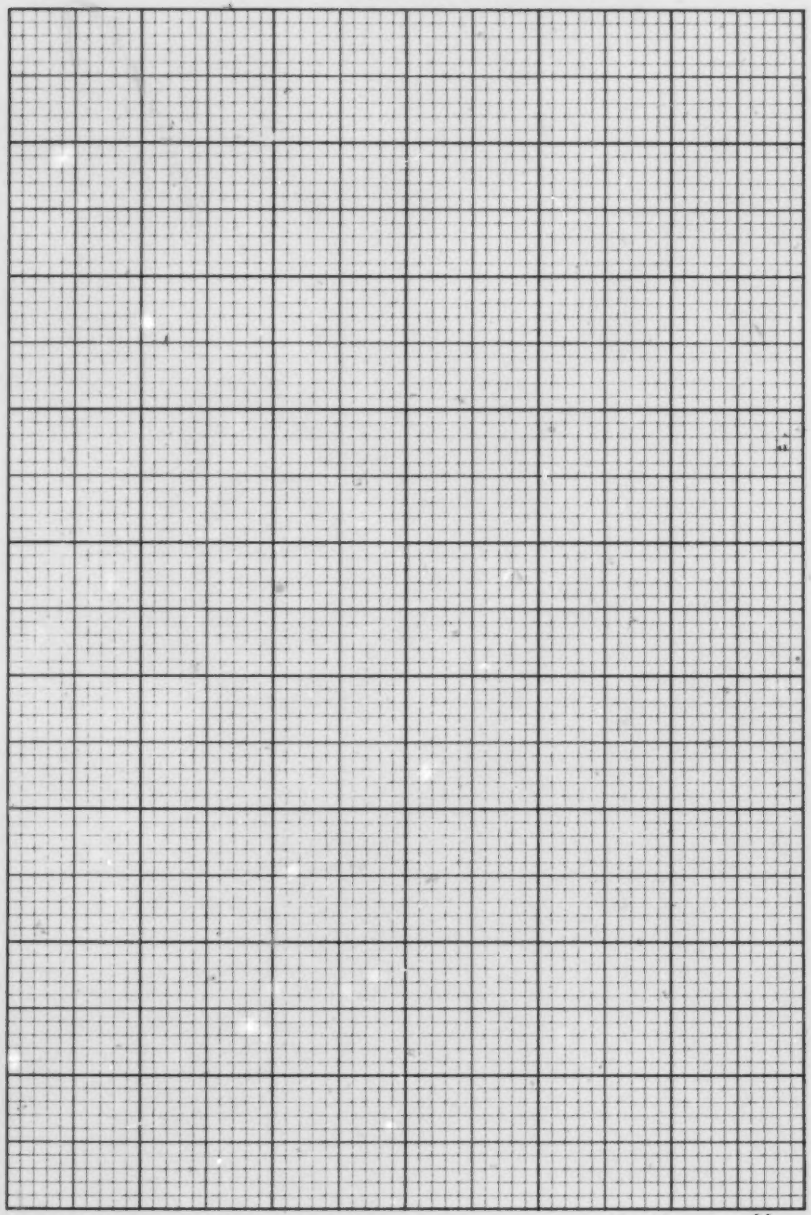
p	r
0	0
.2	.04
.4	.15
.6	.30
.8	.50
1.0	1.00

4. Find the Gini index of a distribution whose Lorenz curve is given by $r = \frac{1}{2}(p+p^2)$. On the basis of this calculation and your answer to question 3, state which distribution is more nearly equal, the one shown in the table above, or the one given by this equation. Explain.

WORK SPACE



WORK SPACE



3.9 Answers to Exercises

1. a) 0
   b) 1

2. b) .01, e) .33, c) .41, a) .73, d) .86.

3. $g = 1 - 2A_L$.

4. a) $1/3 \approx .3333$
   b) $1/2 = .5$
   c) $3/5 = .6$

5. 1st rectangle: $A_1 = .02(.2) = .004$
   2nd rectangle: $A_2 = .1(.2) = .02$
   3rd rectangle: $A_3 = .24(.2) = .048$
   5th rectangle: $A_5 = .74(.2) = .148$

   $A_L = .004 + .02 + .048 + .088 + .148 = .308.$
   $g = 1 - 2(.308) = .384.$

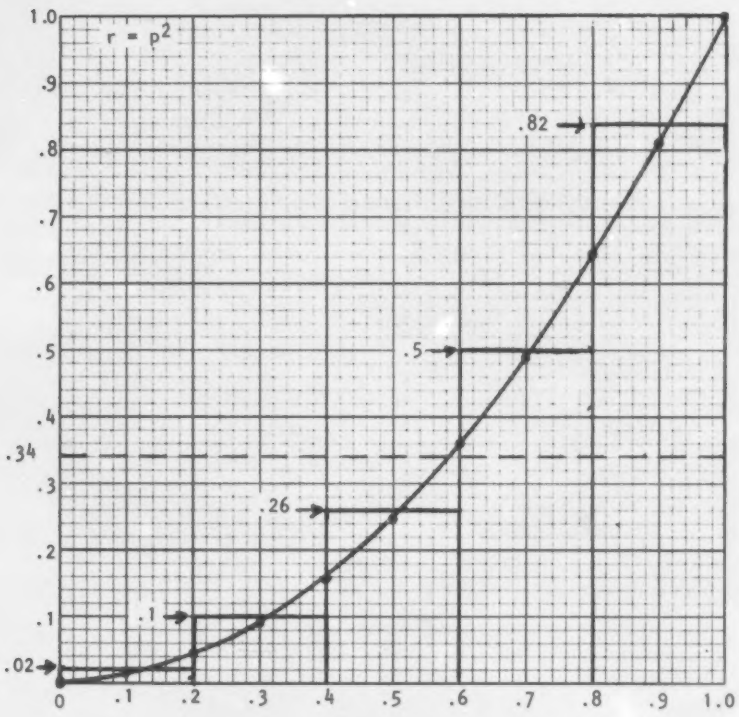
6. Gini index:

Denmark	U.S.	Bolivia
.324	.652	.782

   The distribution of land is most equal in Denmark, next most equal in the U.S., and least equal in Bolivia.

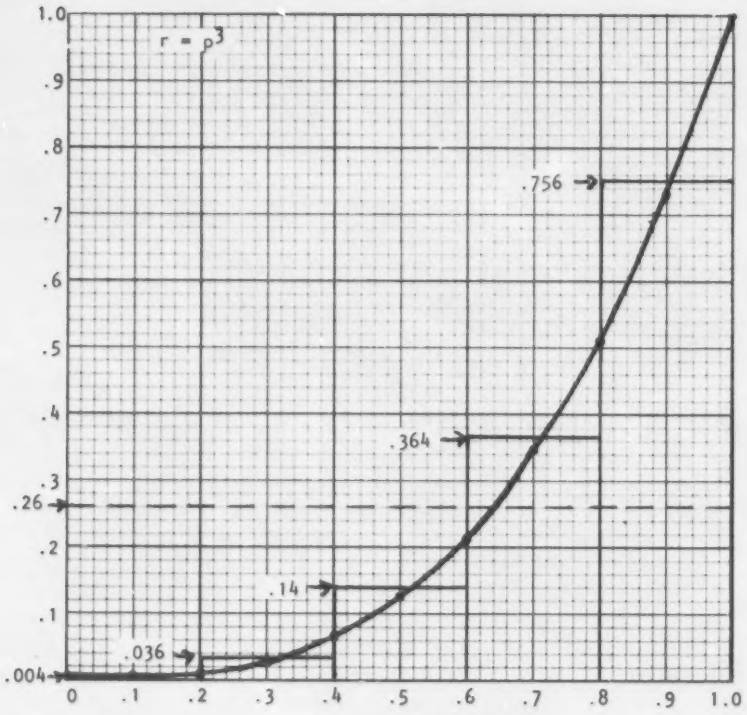
7. $g_{used} \approx .62$
   $g_{produced} \approx .56$

8.



a) Gini index for one rectangle = .32.
b) Gini index for five rectangles = .32.

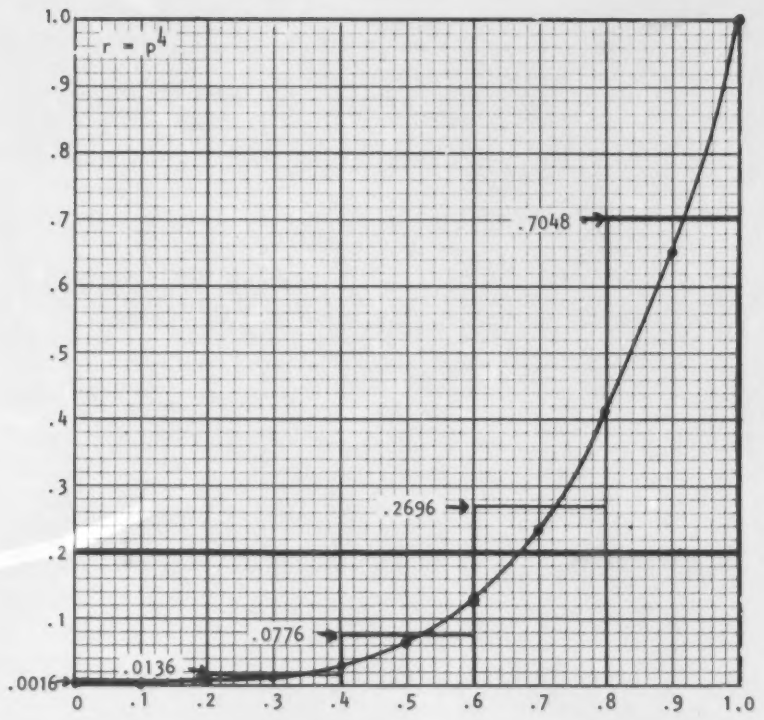
8. (cont.)



a) Gini index for one rectangle = .48.

b) Gini index for five rectangles = .48.

8. (cont.)



a) Gini index for one rectangle = .6.

b) Gini index for five rectangles = .5712.

9. .245, .021

10. Answers should be virtually the same as those obtained in Exercise 7.

11. a) $\int_0^1 p^n \, dp = \frac{p^{n+1}}{n+1}\Big|_0^1 = \frac{1}{n+1}$.

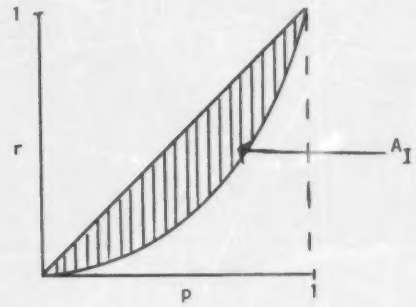
$$g = 1 - 2\left(\frac{1}{n+1}\right) = \frac{n-1}{n+1}.$$

b) for $n = 1$, $g = \frac{1-1}{1+1} = \frac{0}{2} = 0$.

c) $\lim_{n\to\infty} \frac{n-1}{n+1} = \lim_{n\to\infty} \frac{1 - \frac{1}{n}}{1 + \frac{1}{n}} = \frac{1-0}{1+0} = 1.$

3.10 Answers to Model Exam

1. The Gini index is the ratio of the area of inequality to the largest possible area of inequality, i.e., $g = 2A_I$. The smaller the index the more equal the distribution. The larger the index the less equal the distribution.



2. b,c,c,b.

3. .404

4. $\frac{1}{6} = .1667$. The distribution given by $r = \frac{1}{2}(p + p^2)$ is more nearly equal than the distribution of question 4, because it has a smaller Gini index.

The Project would like to thank Judith M. Elkins of Sweet Briar College, John Brevit of Western Kentucky University, Wesley J. Rozema of Northern Arizona University, R. Bruce Mericle of Mankato State University, G. Robert Boynton of the University of Iowa, Thomas W. Casstevens of Oakland University, Spencer E. Minear of Mount Union College, and Karl Schmid of the University of Wisconsin for their reviews, and all others who assisted in the production of this unit.

This unit was field-tested and/or student reviewed in preliminary form at The College of Charleston, Charleston, South Carolina, Humboldt State University, Arcata, California, Fort Lewis College, Durango, Colorado, and Nichols College, Dudley, Massachusetts, and has been revised on the basis of the data received from these sites.

SE036 474

UMAP

MODULES AND MONOGRAPHS IN UNDERGRADUATE MATHEMATICS AND ITS APPLICATIONS

Birkhäuser Boston Inc.
380 Green Street
Cambridge, MA 02139

MODULE 67

Modeling the Nervous System: Reaction Time and the Central Nervous System

by Brindell Horelick, Sinan Koont and Sheldon F. Gottlieb

Biological Applications of Calculus

Intermodular Description Sheet: UMAP Unit 67

Title: MODELING THE NERVOUS SYSTEM: REACTION TIME AND THE CENTRAL NERVOUS SYSTEM

Authors: Brindell Horelick
Dept. of Mathematics
University of Maryland
Baltimore, MD 21228

Sinan Koont
Dept. of Economics
University of Massachusetts
Amherst, MA 01003

Sheldon F. Gottlieb
Dept. of Biology
University of South Alabama
Mobile, AL 36688

Review Stage/Date: IV 8/30/80

Classification: BIO APPL CALC/REACTION TIME-CNS

Prerequisite Skills:
1. Know that the derivative represents a rate of change.
2. Be able to integrate $\int_0^t \frac{u'(t)}{u(t)}\,dt$.
3. Be able to compute $\lim [(At + B)/(Ct + D)]$ as $t \to \infty$.
4. Know the signigicance of $f'(t)$ and $f''(t)$ in the graph of $f(t)$.
5. Be able to differentiate $\ln f(x)$ with respect to x.

This unit is intended for calculus students with an active interest in and some background knowledge of biology. This background knowledge may be represented by concurrent registration in a college level introductory biology course.

Output Skills:
1. Be able to describe the process by which the central nervous system responds to a stimulus.
2. Be able to explain how this process is modeled mathematically, and what conclusions can be drawn from the model regarding reaction time.
3. Be able to discuss the merits of various assumptions concerning the relation between intensity of excitation and stimulus intensity.

Related Units:
Population Growth and the Logistic Curve *(Unit 68)*
The Digestive Process of Sheep *(Unit 69)*
Selection in Genetics *(Unit 70)*

The Project would like to thank Brian J. Winkel of Albion College, Albion, Michigan, George Springer of Indiana University, Bloomington, Indiana, and Marjorie Fitting of San Jose State University, San Jose, California, for their reviews and all others who assisted in the propuction of this unit.

This unit was field-tested and/or student reviewed in preliminary form by Julia P. Kennedy of Georgia State University, Atlanta, Georgia; Jonathan Choate of The Groton School, Groton, Massachusetts; and Richard Montgomery of Southern Oregon State College, Ashland, Oregon, and has been revised on the basis of data received from these sites.

This material was prepared with the partial support of National Science Foundation Grant No. SED76-19615 A02. Recommendations expressed are those of the author and do not necessarily reflect the views of the NSF or the copyright holder.

© 1980 EDC/Project UMAP
All rights reserved.

MODELING THE NERVOUS SYSTEM:
REACTION TIME AND THE CENTRAL NERVOUS SYSTEM

by

Brindell Horelick
Department of Mathematics
University of Maryland Baltimore County
Baltimore, MD 21228

Sinan Koont
Department of Economics
University of Massachusetts
Amherst, MA 01003

Sheldon F. Gottlieb
Department of Biology
University of South Alabama
Mobile, AL 36688

TABLE OF CONTENTS

1. MODELS 1

2. THE NERVOUS SYSTEM 1

3. HOW WE REACT TO A STIMULUS 5

4. A MATHEMATICAL MODEL OF REACTION TIME 7

5. THE RELATION BETWEEN THE EXCITATION INTENSITY E AND THE STIMULUS INTENSITY S 11

6. EXPERIMENTAL CONFIRMATION OF THE MODEL 14

7. REFERENCES 17

8. HINTS AND SUGGESTIONS FOR SOME EXERCISES 18

9. ANSWERS TO SOME EXERCISES 19

MODELING THE NERVOUS SYSTEM:
REACTION TIME AND THE CENTRAL NERVOUS SYSTEM

1. MODELS

Mathematical models, in general, and as applied to biological phenomena are used to describe quantitatively the "real" world. One advantage of a model is that it permits prediction of results mathematically and thereby guides the design of experiments.

Models have limitations. They are valid only to the extent that the data and assumptions on which they are based are valid, and they apply only where the assumptions are valid.

With our present state of knowledge certain subject areas, because of their vast complexity, are not readily amenable to a complete mathematical model. However, complexity of a subject should not be a deterrent to constructing models.

The nervous system, especially that of mammals, is exceedingly complex and does not readily lend itself to a simple model; that does not mean that a model of some aspects of the functioning of the nervous system should not be undertaken.

We shall begin by summarizing some generally accepted background material on the nervous system.

2. THE NERVOUS SYSTEM

2.1 The Nature of the Nervous System

The nervous system, a unique product of evolution, has the ability to respond to a variety of stimuli by generating electrical impulses and transmitting them through nerve cells. Nerve cells vary greatly in length,

from a few micrometers to more than a meter. Nerve cells that transmit impulses from sense organs to the spinal cord and brain are called *sensory neurons* or *afferent neurons*. Nerve cells that convey impulses away from the spinal cord and brain to muscles and glands are *motor neurons* or *efferent neurons*. The neurons in the spinal cord and brain make up the central nervous system (CNS), whereas the neurons outside these structures make up the peripheral nervous system (PNS). [*S-1*]

A neuron is a single cell. A group of neurons tied together forms a *nerve* or nerve bundle (for example, the sciatic nerve). Nerves contain both afferent and efferent neurons. Some neurons are surrounded by a fatty myelin sheath, which is interrupted at 1.0 mm intervals to from the so-called nodes of Ranvier. These nodes are evolutionary adaptations which increase the speed of nerve impulse conduction and conserve energy.

2.2 Definitions

Excitability (irritability) is the ability to react to a stimulus; conductivity is the ability to transmit a state of excitation.

A stimulus - a change in the environment capable of producing a reaction in living tissue (without destroying it) - has a number of attributes that can be quantified. The two most important for our purposes are intensity and duration. Duration is the length of time a stimulus is applied. Intensity may be classified as subthreshold or threshold. A *subthreshold* stimulus is one inadequate to elicit an action potential, but adequate to induce chemical or electrical changes in the nerve at the point of application. A *threshold* stimulus is just strong enough to elicit an action potential in the most excitable neuron in a nerve.

2.3 The Relation Between Strength and Duration

Intuitively one may suspect that the length of time a stimulus has to be applied to elicit an action potential will be a decreasing function of the strength of the stimulus. Indeed, experimental results* confirm that it is reasonable to assume that the strength S of a stimulus is inversely proportional to its duration d. That is, $S = k/d$, where S is measured in milliamperes or in millivolts, t is measured in milliseconds, and k is a constant of proportionality. Thus, there is no single threshold stimulus, but a theoretical range of threshold stimuli of different duration. [*S-2*]

2.4 Neuronal Excitability, Conductivity, and Synaptic Phenomena

The magnitude of response of a neuron is independent of the strength of the stimulus, once threshold level has been reached (the all-or-none law). The neuron responds by electrically conducting the state of excitation, the action potential, down the axon. This conduction process extends from the neuronal cell body, which is uniquely adapted for conducting impulses. [*S-3*]

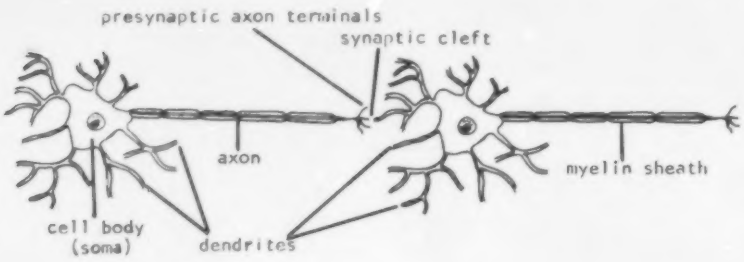


Figure 1. Schematic diagram of two adjacent neurons.

---

* See, for example, W.A.H. Rushton, Journal of Physiology 77: 345, 1933 and J. Erlanger and H.S. Gasser's Electrical Signs of Nervous Activity, University of Pennsylvania Press, Philadelphia, 1937.

Large diameter neurons have a higher excitability than smaller diameter neurons and conduct impulses at a faster rate.

The point of communication that axons make either with other neurons or with muscle cells is termed a synapse. (See Fig. 1.) The axon does not come in direct contact with the other cell, but is separated from it by a minute space, the synaptic cleft. Conduction across a synapse occurs primarily by chemical rather than electrical means. Chemical conduction involves the release of a neurotransmitter substance by the presynaptic axon, its diffusion across the cleft, binding by a receptor on the postsynaptic cell, and the resulting excitation or inhibition of the postsynaptic cell. (See Fig. 2.) Neurotransmitters, especially in the CNS may be inhibitory or excitatory. [S-4]

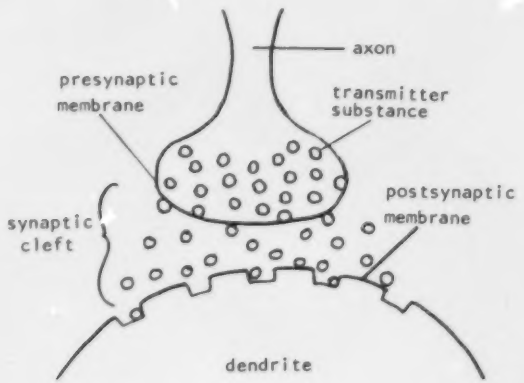


Figure 2. Schematic diagram of synapse.

Hundreds of axons, each from a different neuron, may synapse on a single neuron. Some of these neurons release inhibitory transmitters, others excitatory. The resulting instantaneous state of relative excitation or inhibition of the postsynaptic neuron is dependent upon the temporal and spatial summation of all these influences at any given time.

2.5 The Nerve Message

The basic nerve message consists of the action potential. The message may be altered by the neuron either increasing or decreasing its rate of firing. This latter point is of marked importance since, in the efferent pathways, the rate of firing helps determine the degree of tension developed by a muscle; on the afferent side, the number of impulses is logarithmically related to the stimulus intensity, and influences the excitability of integrative and effector neurons in the CNS.

There are three other means whereby the nerve message may be altered. 1) The intensity of the stimulus, related to the number of neurons excited by the stimulus; 2) the number of branches arising from an axon (axons, near their termination, branch and may synapse with a large number of other neurons or muscle cells); and 3) the synchronous or asynchronous firing of multiple neurons, especially those going to the same effector organ.

The brief introduction above to the nervous system should provide you with a glimpse of its complexity.

3. HOW WE REACT TO A STIMULUS

3.1 Reaction Time

A man inadvertently sticks his finger into an electric outlet and gets a shock. He withdraws his finger quickly (although, alas, not quickly enough). The short, but definitely non-zero, interval between the time he first touches the outlet and the time he begins to withdraw his finger is called his reaction time.

Our intuition, and probably also our experience, suggest that, all other things being equal, the reaction time is probably a decreasing function of the strength of the stimulus. That is, a more intense stimulus results in a faster reaction (a decreased reaction time).

Exactly how does the reaction time depend upon the stimulus intensity? Is it possible to predict this dependence with a mathematical model?

3.2 Dr. Nicholas Rashevsky

To answer these questions, we shall survey some of the work done by Dr. Nicholas Rashevsky of the University of Chicago in an attempt to develop a mathematical model for the central nervous system. Dr. Rashevsky has been a twentieth century pioneer in the development of mathematical biology (and, more specifically, mathematical biophysics). Although we shall only discuss reaction times, Dr. Rashevsky's model also applies to and helps explain many other phenomena.

We shall begin by summarizing some generally accepted background material on the central nervous system, then discuss Rashevsky's model and some results predicted by it, and finally compare these predicted results with experimental data.

3.3 The Afferent Pathway

When an above-threshold stimulus of constant intensity S is applied to a portion of the body, its effect is transmitted to the brain along afferent fibers. The *intensity of excitation* $E = E(S)$ in the afferent pathway is an increasing function of S. It does not change with time. It can be thought of as representing the number of neurons in the afferent pathway excited by the stimulus.

3.4 The Excitatory Factor in the Brain

This excitation in the afferent pathway prompts a response within the brain. The exact physiological nature of this response is not clear, but it is known to build up with the passage of time (assuming the stimulus continues). We may think of it, for example, as the creation of a central excitatory state, with increasing frequency of excitation of these neurons within the brain. Whatever the nature of the response, we shall use $\varepsilon(t)$, to denote

the magnitude of this excitatory state at time t. We call $\varepsilon(t)$ the *excitatory factor*.

3.5 The Efferent Pathway

Whatever the physiological description of $\varepsilon(t)$, when (and if) it builds up to a certain threshold level h_E it stimulates the efferent neuron, leading from the brain to an effector organ. This results in a response to the stimulus. Now for the computations.

4. A MATHEMATICAL MODEL OF REACTION TIME

4.1 Three Time Intervals

The reaction time T is the sum of three times:

(1) $$T = t_A + t_B + t_E$$

(See Figure 3) where

(a) t_A is the time required for the excitation in the afferent pathway to be transmitted from the point of application of the stimulus to the brain;

(b) t_B is the time required for $\varepsilon(t)$ to grow to the threshold level h_E;

(c) t_E is the time required for the response to be transmitted along the efferent pathway.

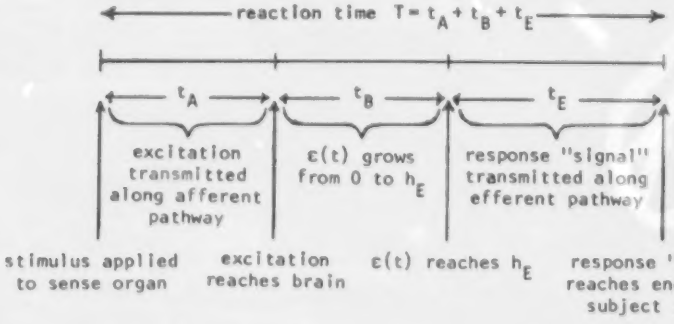

reaction time $T = t_A + t_B + t_E$
t_A
t_B
t_E
excitation transmitted along afferent pathway
$\varepsilon(t)$ grows from 0 to h_E
response "signal" transmitted along efferent pathway
stimulus applied to sense organ
excitation reaches brain
$\varepsilon(t)$ reaches h_E
response "signal" reaches end organ; subject reacts

Figure 3. How Central Nervous System Responds to a Stimulus (Schematic Diagram)

Of these, only t_B is dependent upon the stimulus intensity S. Remember that the excitation intensity E depends on S, but its transmission time t_A does not. Also, the intensity of excitation in the efferent pathway may conceivably depend upon S -- that does not concern us here -- but its transmission time t_E does not depend on S.

4.2 Identifying the Problem

The heart of our problem is to determine t_B as a function of S. But all that we know about t_B is that it is a certain time interval; namely, the interval from when $\varepsilon(t)$ starts growing to when $\varepsilon(t) = h_E$. So we need to know something about the function $\varepsilon(t)$.

Since we are interested in a certain time interval, we shall call the beginning of this interval $t = 0$. Then the end of the interval occurs at $t = t_B$, since its length is t_B. In other words, t_B is the solution of

(2) $$\varepsilon(t_B) = h_E.$$

Since the afferent pathway excitation reaches the brain t_A seconds after the initial application of the stimulus, the stimulus was applied at time $t = -t_A$. See Figure 4.

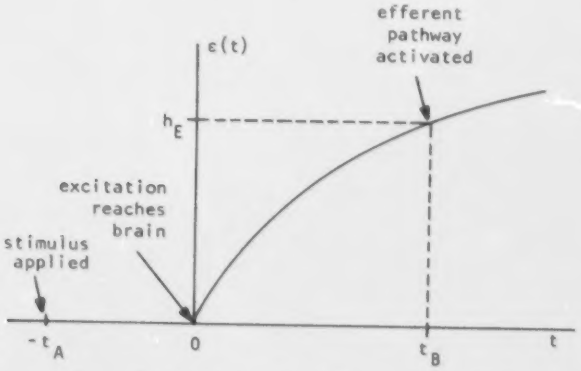


Figure 4. Growth of excitatory factor $\varepsilon(t)$.

4.3 Assumptions about the Excitatory Factor $\varepsilon(t)$

We shall make two critical assumptions about $\varepsilon(t)$. Both are well supported by biological theory.

(A) If each excited nerve fiber in the afferent pathway does its part in exciting neuron cycles, producing the central excitatory state, or whatever, the rate at which this response is excited is proportional to E.

(B) There is nothing counteracting the tendency for $\varepsilon(t)$ to increase other than natural "decay" (for example, accidental failure of a neuron cycle, somewhat like a light bulb's burning out). This decay occurs randomly (for example, any neuron cycle has the same chance of failing as any other), and the amount of decay per second is proportional to the amount of excited neuron cycles, chemical, or whatever present.

Mathematically these two assumptions say

$$\varepsilon'(t) = k_1 E - k_2 \varepsilon(t) \tag{3}$$

where k_1 and k_2 are in positive constants.

4.4 Computation of T in Terms of E

Equation (3) enables us to get expressions for $\varepsilon(t)$, t_B, and T in terms of E. We shall do this, and then consider how to express E in terms of S.

We can rewrite equation (3)

$$\frac{\varepsilon'(t)}{k_1 E - k_2 \varepsilon(t)} = 1$$

and then antidifferentiate:

$$\int \frac{\varepsilon'(t)}{k_1 E - k_2 \varepsilon(t)}\, dt = \int 1\, dt.$$

Putting $u(t) = k_1 E - k_2 \varepsilon(t)$, so $u'(t) = -k_2 \varepsilon'(t)$,

we get $$-\frac{1}{k_2}\int \frac{u'(t)}{u(t)}\, dt = \int 1\, dt,$$

$$-\frac{1}{k_2} \ln u(t) = t + C. \tag{4}$$

There is no need for an absolute value sign because u(t) is positive throughout the interval of integration.

Putting $t = 0$ and using the fact that

$$u(0) = k_1E - k_2\varepsilon(0) = k_1E,$$

we get, from equation (4),

$$C = -\frac{1}{k_2} \ln (k_1E).$$

Then, using equation (4) again,

$$t = -\frac{1}{k_2} \ln u(t) + \frac{1}{k_2} \ln (k_1E),$$

and, finally,

$$t = -\frac{1}{k_2}[\ln (k_1E - k_2\varepsilon(t)) - \ln (k_1E)].$$

$$t = \frac{1}{k_2} \ln \frac{k_1E}{k_1E - k_2\varepsilon(t)}. \tag{5}$$

Now placing $t = t_B$ and using equation (2),

$$t_B = \frac{1}{k_2} \ln \frac{k_1E}{k_1E - k_2\varepsilon(t_B)}$$

$$t_B = \frac{1}{k_2} \ln \frac{k_1E}{k_1E - k_2h_E}.$$

From equation (1) we get immediately

$$T = t_A = t_E + \frac{1}{k_2} \ln \frac{k_1E}{k_1E - k_2h_E}$$

or, writing $T_0 = t_A + t_E$,

$$T = T_0 + \frac{1}{k_2} \ln \frac{k_1E}{k_1E - k_2h_E}. \tag{6}$$

Exercise 1. (a) Solve equation (5) explicitly for $\varepsilon(t)$.

(b) Prove $\varepsilon(t)$ is an increasing function.

(c) Compute $\lim_{t\to\infty} \varepsilon(t)$, and show that it exceeds h_E if and only if $k_2 h_E < k_1 E$.

(d) Determine how large E must be if a response is to take place.

(e) Use parts (b) and (c) to sketch a graph of $\varepsilon(t)$.

Exercise 2. Your "common sense" might suggest that an increase in the excitation intensity E results in a decrease in the reaction time T. Show that the model predicts this by considering T as a function of E and showing $T'(E) < 0$ whenever $E > k_2 h_E / k_1$. (This last condition is necessary because, according to Exercise 1, T(E) is not even defined when $E < k_2 h_E / k_1$.)

Exercise 3. Assume that a stimulus has been applied long enough for $\varepsilon(t)$ to be essentially constant ($\varepsilon(t) \simeq \lim_{t\to\infty} \varepsilon(t)$). At time $t = t^* - t_A$ the stimulus is cut off (so that E drops to 0 at the brain at time $t = t^*$). Obtain an expression for $\varepsilon(t)$ for $t \geq t^*$.

5. THE RELATION BETWEEN THE EXCITATION INTENSITY E AND THE STIMULUS INTENSITY S

What can we say about E as a function of S? Very little, actually, on the basis of direct observation or developed biological theory. What we can do is:

(1) use our biological knowledge to predict the "shape" of the graph of E(S);

(2) write down a graph whose function has this "shape";

(3) plug this function into equation (6) and check the resulting values of T for various values of S against empirical data.

To illustrate this approach we shall consider three different submaximal stimuli.

5.1 Weak Stimuli

We have already suggested an interpretation of E in terms of the number of excited neurons in the afferent pathway. These neurons, like those in the efferent pathway, have a threshold level h_A below which no excitation occurs. So if $S < h_A$, then $E(S) = 0$. We can say, then, that E is entirely accounted for by the amount by which S exceeds h_A, that is, by $S - h_A$. The simplest possible assumption along these lines is proportionality:

$$E(S) = k_3(S - h_A) \qquad \text{for } S > h_A \tag{7}$$

where k_3 is a positive constant. Applying equation (7) to equation (6) gives

$$T = T_0 + \frac{1}{k_2} \ln \frac{S - h_A}{S - h_A - C} \tag{8}$$

where $C = k_2 h_E / k_1 k_3$. We'll leave the details to you (Exercise 4).

5.2 Moderate Stimuli

Equation (7) works well for weak stimuli. But it cannot be exactly right if, as we have suggested, E measures the number of excited neurons in the afferent pathway. After all, S can increase without bound, but there is only a finite number of neurons.

In equation (7), $E'(S)$ is a positive constant k_3. But as the stimulus grows stronger, the number of available (not yet excited) neurons decreases. This suggests that fewer additional ones are excited per unit increase in S. In other words, $E'(S)$ is still positive but growing smaller; $E''(S) < 0$. In graphical terms, the curve E(S) will begin to level off (see Figure 5). Perhaps the simplest function having this shape and satisfying $E(h_A) = 0$ is

(9) $$E(S) = k_3 \ln \frac{S}{h_A} \qquad \text{for } S > h_A$$

where, as before, k_3 is a positive constant.

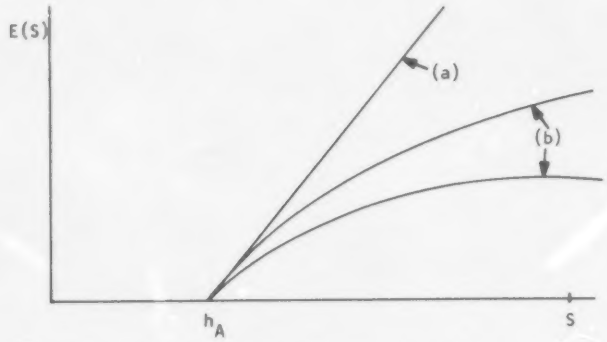


Figure 5. (a) Sketch of $E_w(S) = k_3(S-h_A)$ $[E_w(h_A)=0;\ E_w'(S)=k_3]$
(b) Sketches of hypothetical curves E(S) with $E(h_A)=0$, $E'(h_A)=k_3$, $E'(S) > 0$, and $E''(S) < 0$.

Applying equation (9) to equation (6) gives

(10) $$T = T_0 + \frac{1}{k_2} \ln \frac{\ln S - \ln h_A}{\ln S - \ln h_A - C} .$$

Again, $C = k_2 h_E / k_1 k_3$, and again we'll leave the details to you (Exercise 5).

5.3 Strong Stimuli

For sufficiently strong stimuli, even equation (9) is unsatisfactory. After all, $\lim_{S\to\infty} E(S) = \infty$, while, as we have already indicated, there is an upper bound on the excitation, corresponding to *all* the neurons in the afferent pathway being excited. In this case a function often used is

(11) $$E(S) = k_3 \ln \frac{S}{h_A + Se^{-E^*/k_3}}$$

where k_3 and E^* are positive constants. In Exercise 4 you will be asked to investigate why this may be a better approximation.

Exercise 4. Derive equation (8) from equations (6) and (7). Remember that $C = k_2 h_E / k_1 k_3$.

Exercise 5. Derive equation (10) from equations (6) and (9). Again, remember that $C = k_2 h_E / k_1 k_3$.

Exercise 6. If E(S) is defined by equation (11):

(a) Compute $\lim_{S \to \infty} E(S)$.

(b) Show that $E'(S) > 0$ for all $S > 0$.

(c) Show that $E''(S) < 0$ for all $S > 0$.

(d) Use (a), (b) and (c) to sketch a graph of E(S).

(e) Discuss in what ways equation (11) may give a better approximation of the relation between E and S than either equation (7) or equation (9).

---

6. EXPERIMENTAL CONFIRMATION OF THE MODEL

In view of the many more or less arbitrary assumptions along the way, it may come as a surprise to you that equations (8) and (10) do a good job of representing experimental data for low and moderate stimuli respectively. Figures 6 and 7 show T plotted agains $\log_{10} S$. The experimental data are for visual and auditory stimuli respectively. The solid curves are graphs of equation (10), where in each case the constants were chosen so that the curve fit the experimental data as well as possible. Figure 8 shows T plotted against S for taste stimuli. The solid curve there is obtained from equation (8).

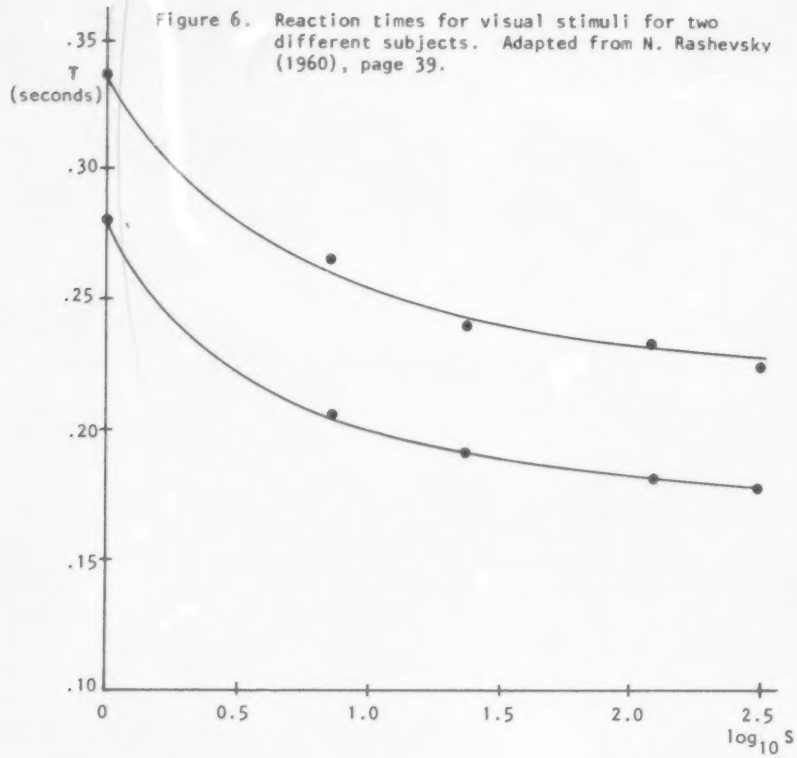


Figure 6. Reaction times for visual stimuli for two different subjects. Adapted from N. Rashevsky (1960), page 39.

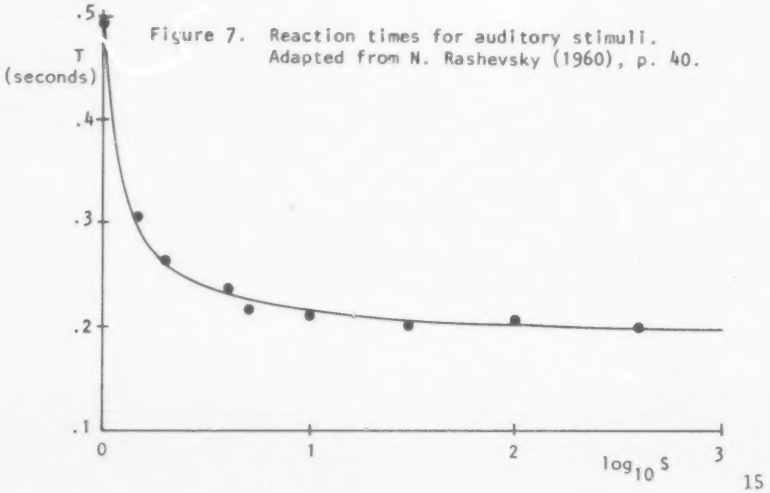


Figure 7. Reaction times for auditory stimuli. Adapted from N. Rashevsky (1960), p. 40.

15

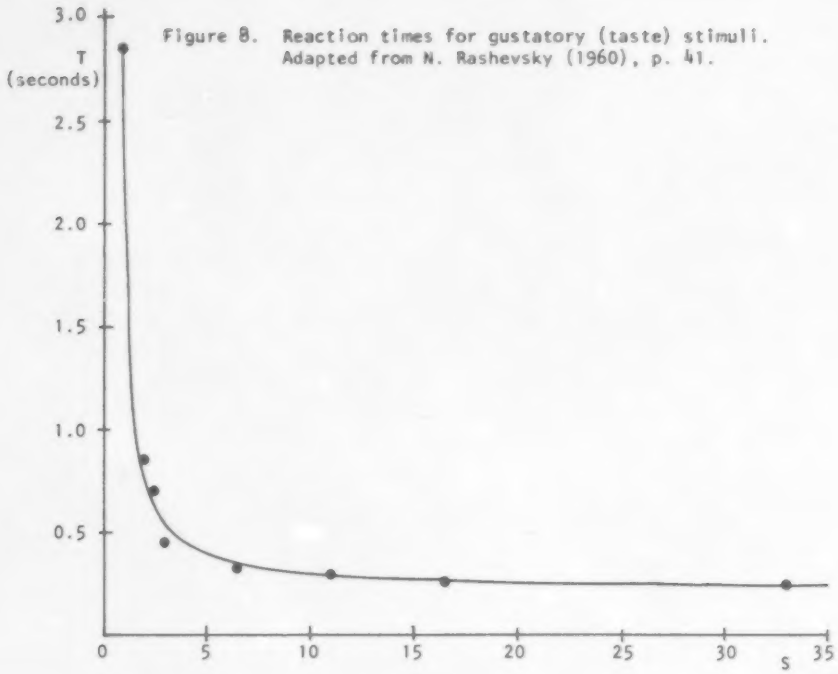


Figure 8. Reaction times for gustatory (taste) stimuli. Adapted from N. Rashevsky (1960), p. 41.

WORK SPACE

16

Exercise 7. (A calculator will be helpful.) The upper curve in Figure 6 was obtained from equation (10) with $T_0 = 0.195$, $k_2 = 2.421$, $h_A = 0.141$, and $C = 0.57$. Confirm that the graph of this curve is accurate by computing points on the curve corresponding to $S = 1$, 7, 23, 120, and 310 (the stimuli of the original experiment).

BE CAREFUL: In Figure 6, T is plotted against $\log_{10} S$, but in equation (10), the formula for T uses ln S.

Exercise 8. (A calculator will be helpful.) Do the same as in Exercise 7 for the lower curve of Figure 6. Use $T_0 = 0.155$, $k_2 = 3.584$, $h_A = 0.208$, and $C = 0.57$.

Exercise 9. (A calculator will be helpful.) Do the same as in Exercise 7 for the curve of Figure 7. This time $T_0 = 0.187$, $k_2 = 3.289$, $h_A = 0.691$, and $C = 0.23$. The stimuli used in the experiment were $S = 1$, 1.5, 2, 4, 5, 10, 30, 100, and 400.

Exercise 10. (A calculator will be helpful.) The curve in Figure 8 was obtained from equation (8) with $T_0 = 0.22$, $k_2 = 0.263$, $h_A = 0.6$, and $C = 0.2$. Confirm that the graph of this curve is accurate by computing points on the curve corresponding to $S = 1$, 2, 2.5, 3, 6.5, 11, 16.5, and 33 (the stimuli of the original experiment).

7. REFERENCES

Curtis, H. (1975). Biology, second edition, Worth Publishers New York.

Rashevsky, N. (1960). Mathematical Biophysics, v. 2, Dover, New York.

Rashevsky, N. (1961). Mathematical Principles in Biology and Their Applications, C.C. Thomas, Springfield, Illinois.

8. HINTS AND SUGGESTIONS FOR SOME EXERCISES

1. (a) Multiply both sides of equation (5) by k_2, and then exponentiate (i.e., take antilogs base e) both sides, using the fact that $e^{\ln x} = x$. Then clear of fractions, etc.

   (b) Differentiate the answer to (a) with respect to t.

   (c) Start with the answer to (a), and recall that

   $$\lim_{t \to \infty} e^{-k_2 t} = 0$$

   since $k_2 > 0$.

   (d) For a response to occur, we must have $\varepsilon(t) = h_E$ for some t. See parts (b) and (c).

   (e) You know the graph of $\varepsilon(t)$ is constantly rising (why?) and approaches $y = k_1E/k_2$ (why?). Do not attempt point by point plotting.

2. Use equation (6). It helps to write

   $$\ln \frac{k_1E}{k_1E - k_2h_E} = \ln k_1E - \ln (k_1E - k_2h_E).$$

3. Equation (3) applies, except that now $E = 0$. So

   $$\frac{\varepsilon'(t)}{\varepsilon(t)} = -k_2.$$

   Follow Section 4.4 (except that this is easier). To find C, use $\varepsilon(t^*) = k_1E/k_2$ (why?).

6. (a) Use the fact that

   $$\lim_{s \to \infty} \frac{S}{m + nS} = \frac{1}{n},$$

   and the fact that if $f(S) \to L$ then $\ln f(S) \to \ln L$.

(b) Differentiate E(S) and show the result can be written

$$h_A k_3 \, (Sh_A + S^2 e^{-E^*/k_3})^{-1},$$

and that this expression is always positive.

(c) Write E'(S) as in hint 6(b) and differentiate. Remember that $m^{-2} > 0$ for all $m \neq 0$.

7. To compute T(S) for various values of S, first compute $\ln h_A = \ln 0.141 = -1.96$, $\ln h_A + C = -1.96 + 0.57 = -1.39$, and $\frac{1}{k_2} = \frac{1}{2.421} = 0.413$. Then equation (9) becomes

$$T(S) = 0.195 + 0.413 \ln \frac{\ln S + 1.96}{\ln S + 1.39}.$$

8. Proceeding as in the hint to Exercise 7, we get

$$T(S) = 0.155 + 0.279 \ln \frac{\ln S + 1.57}{\ln S + 1.00}.$$

9. Proceeding as in the hint to Exercise 7, we get

$$T(S) = 0.187 + 0.304 \ln \frac{\ln S + 0.37}{\ln S + 0.14}.$$

10. Since $\frac{1}{k_2} = \frac{1}{0.263} = 3.8$, equation (8) becomes

$$T(S) = 0.22 + 3.8 \ln \frac{S - 0.6}{S - 0.8}.$$

9. ANSWERS TO SOME EXERCISES

1. (a) $c(t) = \frac{k_1 E}{k_2} \cdot \frac{e^{k_2 t} - 1}{e^{k_2 t}} = \frac{k_1 E}{k_2} (1 - e^{-k_2 t})$

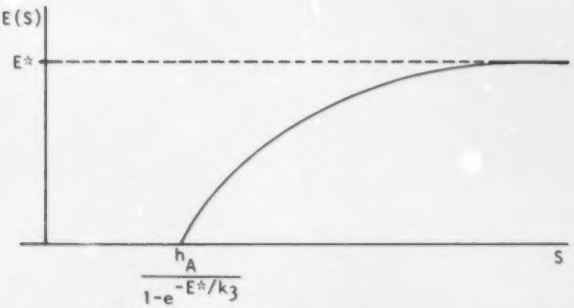
(c) $\frac{k_1 E}{k_2}$

(d) $E > \frac{k_2 h E}{k_1}$

3. $\varepsilon(t) = \frac{k_1 E}{k_2} e^{k_2(t^* - t)} = \frac{k_1 E}{k_2} e^{-k_2(t-t^*)}$

4. (a) E^*

(d)



(e) Equation (10) is the only one of the three proposed equations for which E(S) has an upper bound ($\lim_{S\to\infty} E(S) \neq \infty$).

5.

S	1	7	23	120	310
$\log_{10}$ S	0	0.85	.36	2.08	2.49
T	.337	.260	.244	.231	.227

6.

S	1	7	23	120	310
$\log_{10}$ S	0	0.85	1.36	2.08	2.49
T	.281	.204	.191	.181	.178

7.

S	1	1.5	2	4	5	10	30	100	400
$\log_{10}$ S	0	0.18	0.30	0.60	0.70	1.00	1.48	2.	2.60
T	.482	.294	.261	.230	.225	.214	.206	.2	.198

8.

S	1	2	2.5	3	6.5	11	16.5	33
T	2.85	0.81	0.64	0.55	0.35	0.29	0.27	0.24

UMAP

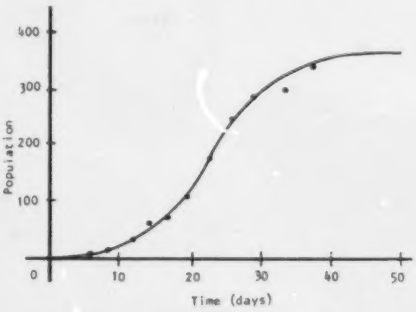
MODULES AND MONOGRAPHS IN UNDERGRADUATE MATHEMATICS AND ITS APPLICATIONS

Birkhäuser Boston Inc.
380 Green Street
Cambridge, MA 02139

MODULE 68

Population Growth and the Logistic Curve

by Brindell Horelick
Sinan Koont
and
Sheldon F. Gottlieb



Biological Applications of Calculus

SE036 474

POPULATION GROWTH AND THE LOGISTIC CURVE

by

Brindell Horelick
Department of Mathematics
University of Maryland Baltimore County
Baltimore, Maryland 21228

Sinan Koont
Department of Economics
University of Massachusetts
Amherst, Massachusetts 01003

and

Sheldon F. Gottlieb
Indiana University - Purdue University
Fort Wayne, Indiana 46805

TABLE OF CONTENTS

1. GROWTH 1

2. THE MODEL 2

3. THE LOGISTIC EQUATION 6

4. EXPERIMENTAL VERIFICATION 8

5. DETERMINING MATHEMATICALLY WHETHER GROWTH IS LOGISTIC . . . 10

6. CRITICISM OF THE MODEL 16

7. POSTSCRIPT 17

8. REFERENCES 17

Intermodular Description Sheet: UMAP Unit 68

Title: POPULATION GROWTH AND THE LOGISTIC CURVE

Author: Brindell Horelick
Department of Mathematics
University of Maryland Baltimore County
Baltimore, Maryland 21228

Sinan Koont
Department of Economics
University of Massachusetts
Amherst, Massachusetts 01003

Sheldon F. Gottlieb
Indiana University - Purdue University
Fort Wayne, Indiana 46805

Review Stage/Date: IV 7/17/79

Classification: BIO APPL CALC/POPULATION GROWTH

Suggested Support Material: A calculator, or a table of natural logarithms, and a transparent ruler or straightedge.

Prerequisite Skills:

1. Know that $f'(t)$ describes the rate of change of $f(t)$, and be able to integrate $\int_0^t \frac{u'(t)}{u(t)}\,dt$.
2. Know the graphical significance of $f'(t)$ and $f''(t)$ and be able to find extreme points and points of inflection of $f(t)$ in simple cases.
3. An acquaintance with the functions $\ln x$ and e^x.

Output Skills:

1. Be able to describe the mathematical assumptions that lead to the logistic equation.
2. Be able to criticize the logistic model, and to discuss some of its strengths and weaknesses.
3. Know the behavior of the logistic function, and the shape of its graph.

Other Related Units:

The following units involve the exponential function, and apply differential functions to biology:

Modeling the Nervous System: Reaction Time and the Central Nervous System *(Unit 67)*
The Digestive Process of Sheep *(Unit 69)*
Selection in Genetics *(Unit 70)*

The following monograph is also available:

Introduction to Population Modeling by James C. Frauenthal.

This material was prepared with the partial support of National Science Foundation Grant No. SED76-19615 A02. Recommendations expressed are those of the authors and do not necessarily reflect the views of the NSF or the copyright holder.

© 1979 EDC/Project UMAP
All rights reserved.

POPULATION GROWTH AND THE LOGISTIC CURVE

1. GROWTH

Growth, a fundamental characteristic of life, is a term that can be applied to an individual or to a population. For individuals, growth can be viewed as an irreversible increase in the size of an organism or as a permanent increase in its mass (the proportional increase in the chemical constituents of the organism), or both. For a population, growth is an increase in the number of individuals.

The way growth is measured may vary, depending on the organism and the information required. For example, growth of unicellular organisms is measured in terms of increase in cell mass and cell number. One could measure the linear increase in cell size, but this is a tedious process. For larger organisms, growth can be expressed either as an increase in mass or as a linear increase in size.

The growth of a population is affected by the genetic constitution of organisms within the population, and may be affected by the concentration of available nutrients or the accumulation of toxic metabolic products. It may also be affected by external factors such as decreasing availability of space and changed climatic conditions, or by the presence of predators, including highly infectious and virulent microorganisms.

Most growth curves have the characteristic exponential shape shown in Figure 3 in Section 3. The initial portion of the curve shows a lag phase in which

the growth rate is very slow. As the growth rate and population size gradually increase, they reach a period of rapid increase, the exponential growth phase. Eventually, one or more of the nutrient, toxic, or environmental factors mentioned above causes a rapid deceleration of the growth rate.

Typically, there is a limiting value A beyond which the size of the population cannot increase, since the factors inhibiting growth overcome those contributing to growth as the population approaches this size. Ecologists call this A the *carrying capacity of the environment;* microbiologists call it the *stationary phase.*

In this module, we will model the growth pattern we have been describing mathematically, and obtain an equation which gives the size of the population f(t) as a function of time t. Our analysis begins as follows.

2. THE MODEL

2.1 Two Assumptions

We suppose that f(t) is a differentiable function of t, and make two further assumptions in an attempt to describe mathematically the behavior we have observed the population to exhibit in its early and late stages of growth.

(a) At first, f'(t) increases with f(t), and as long as f(t) remains much smaller than A, the rate f'(t) at which the population increases is approximately proportional to the number f(t) of organisms present. The larger the population at this early stage, the faster it grows.

(b) As $f(t)$ approaches A, it grows more and more slowly. That is, $f'(t)$ gets small as $A-f(t)$ gets small. The larger the population at this later stage, the slower it grows.

2.2 The Effect of Each Assumption

Assumption (a) says mathematically that

$$f'(t) \approx k_1 f(t) \tag{1}$$

or

$$\frac{f'(t)}{f(t)} \approx k_1$$

for $f(t)$ near zero but not equal to zero, where k_1 is a positive constant. Assumption (b) is less precise, since it does not mention proportionality explicitly, but one mathematical realization of it, for $f(t)$ near A but not equal to A, is

$$f'(t) \approx k_2(A-f(t)) \tag{2}$$

or

$$\frac{f'(t)}{A-f(t)} \approx k_2,$$

where k_2 is a positive constant. We leave it to you (Exercises 1 and 2) to show that if we change the approximation signs in Equations (1) and (2) to equal signs and then solve them we get, respectively,

$$f(t) = B_1 e^{k_1 t} \tag{3}$$

and

$$f(t) = A - B_2 e^{-k_2 t}, \tag{4}$$

where B_1 and B_2 are *positive* constants. The graphs of these functions depend, of course, on the values of the constants A, B_1, B_2, k_1, and k_2. Typical graphs for the functions are shown in Figures 1 and 2.

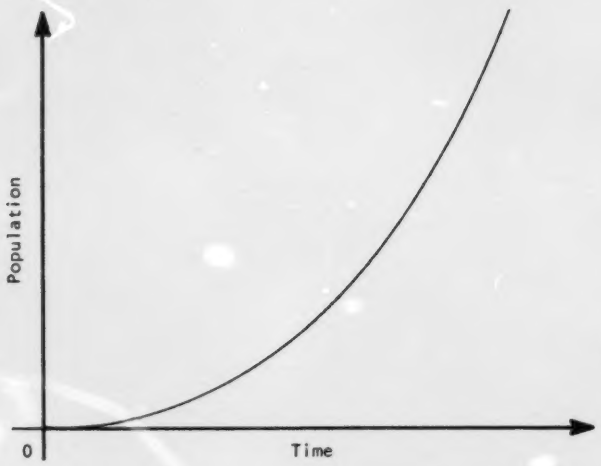


Figure 1. Typical population growth satisfying assumption (a).

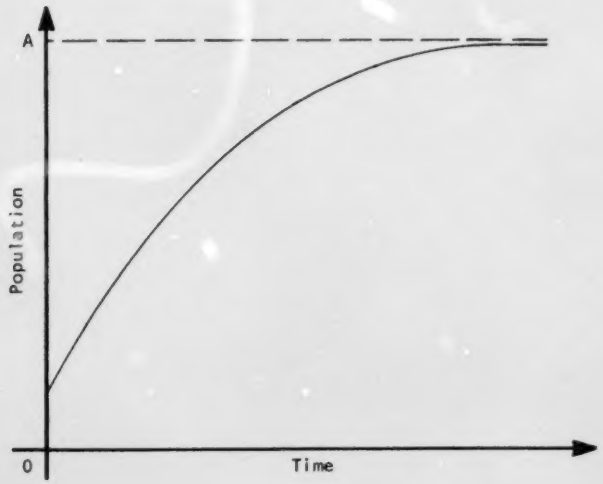


Figure 2. Typical population growth satisfying assumption (b).

Exercises.

1. Let k_1 be a positive constant, and $f(t)$ a differentiable function of t whose values are positive and which satisfies the equation

$$f'(t) = k_1 f(t).$$

Show that

$$f(t) = B_1 e^{k_1 t},$$

where B_1 is a positive constant. Be sure to explain why B_1 must be positive.

2. Let k_2 be a positive constant, and $f(t)$ a differentiable function of t whose values are less than A, and which satisfies the equation

$$f'(t) = k_2(A - f(t)).$$

Show that

$$f(t) = A - B_2 e^{-k_2 t},$$

where B_2 is a positive constant. Be sure to explain why B_2 must be positive.

2.3 Combining the Assumptions

There are many equations that fit both assumptions (a) and (b). Perhaps the simplest of these is

$$f'(t) = kf(t)(A - f(t)), \tag{5}$$

where k is a positive constant. If $f(t)$ is tiny compared to A then

$$A - f(t) \approx A$$

and so

$$f'(t) \approx Akf(t),$$

which is Equation (1) with $k_1 = Ak$. On the other hand, if $A - f(t)$ is tiny compared to A then

$$f(t) = A,$$

and so

$$f'(t) = kA(A-f(t)),$$

which is Equation (2) with $k_2 = kA$.

It seems reasonable to expect that any function f(t) satisfying Equation (5) will have a graph which looks very much like Figure 1 for small f(t), and like Figure 2 for f(t) near A. Let us try to find f(t) explicitly and see if this is so.

Exercises

3. Criticize this model. Name some important considerations which are not taken into account. Suggest some other equation satisfying (a) or (b) or both. Name some organisms for which you think this model might be accurate, and some for which it probably would not be.

---

3. THE LOGISTIC EQUATION

If $f(t) \neq 0$ and $f(t) \neq A$, we can rewrite Equation (5) in the following way:

$$\frac{f'(t)}{f(t)[A-f(t)]} = k. \tag{6}$$

We may further rewrite the left side of Equation 6 by using the partial fraction decomposition

$$\frac{1}{x(A-x)} = \frac{1}{A}\left(\frac{1}{A-x} + \frac{1}{x}\right).$$

When we do so, with $x = f(t)$, Equation (6) becomes

$$\frac{1}{A}\left(\frac{f'(t)}{A-f(t)} + \frac{f'(t)}{f(t)}\right) = k.$$

We then multiply both sides of this equation by -A and antidifferentiate to get

$$\int \left(\frac{-f'(t)}{A-f(t)} - \frac{f'(t)}{f(t)} \right) dt = -Ak \int dt,$$

$$\ln [A-f(t)] - \ln f(t) = -Akt + C,$$

and, finally,

$$\ln \frac{A-f(t)}{f(t)} = -Akt + C.$$

We are able to omit the usual absolute value signs in these computations becuase the values of $f(t)$ and $[A-f(t)]$ cannot be negative in the context of our application. It follows that

$$(7) \qquad \frac{A-f(t)}{f(t)} = e^{-Akt+C}$$

$$= Be^{-Akt},$$

where we have written

$$B = e^{C}.$$

Solving Equation (7) for $f(t)$, we get

$$(8) \qquad f(t) = \frac{A}{1 + Be^{-Akt}}.$$

Equation (8) is known as the *logistic equation.* Its graph (Figure 3) is an S-shaped curve, as predicted in Section 2.3.

Exercises

4. a. Find the point of inflection of the graph of $f(t)$. What is the value of $f(t)$ at this point?

   b. Find the point where $f'(t)$ reaches its maximum.

5. Suppose a logistic equation is given by

$$f(t) = \frac{500}{1+e^{2}e^{-t/10}} \text{ (time in days).}$$

   a. What is the approximate maximum size of the population?

b. What is the approximate initial size of the population?

c. In how many days does the population reach half its maximum size?

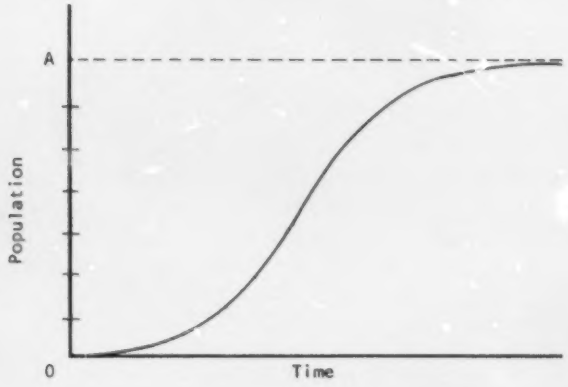


Figure 3. A typical logistic curve. The curve gets the name "logistic" not from biology, but from curves like it used in the early 1700's to explain the properties of logarithms.

4. EXPERIMENTAL VERIFICATION

Despite the rather primitive nature of assumptions (a) and (b), the logistic equation has been found to give an accurate description of the behavior of certain yeast, bacteria, protozoa, and insect population. We cite two examples.

4.1 A *Drosophila* Population

A.J. Lotka (see References) cites the results of a controlled experiment in *Drosophila* population. The dots in Figure 4 represent the data, and the solid curve is a theoretical logistic curve fitted to the data through an appropriate choice of A, B, and k.

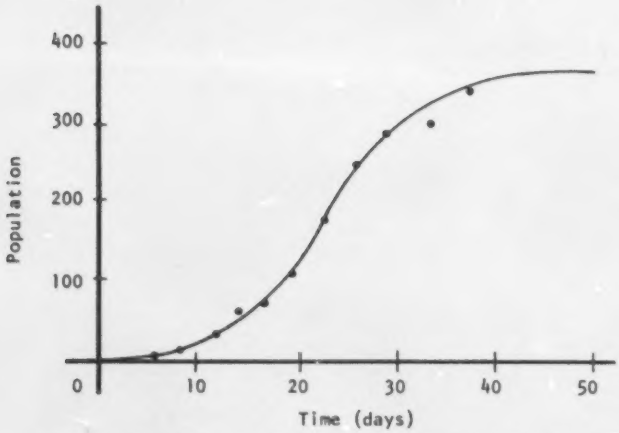


Figure 4. The growth of a *Drosophila* population in a controlled experiment.

(Source: A.J. Lotka, Elements of Mathematical Biology, Dover, New York, 1956, p. 69.)

4.2 A Bacterial Population

The size of a population of bacteria may be expressed as the number of colony-forming units per milliliter. Assuming that the number of bacteria per milliliter is fixed, this simply changes the unit of measurement, and does not affect the applicability of Equation (8).

A bacterial count at the Indiana University-Purdue University department of biological sciences yielded the data in the first two columns of Table 1, which follows. The values in the third colum are computed from the logistic equation.

$$(9) \qquad f(t) = \frac{3.1 \times 10^7}{1 + 22026e^{-0.6262t}} \text{ (time in hours).}$$

TABLE 1

Data from an Experiment
with a Bacteria Population

t (hours)	Bacteria per milliliter	f(t) (from Equation 9)
0	1.2×10^3	1.4×10^3
1	3.9×10^3	
4	1.61×10^4	1.72×10^4
5.5	4.39×10^4	
8	2.83×10^5	2.09×10^5
10	8.90×10^5	
13.5	4.83×10^6	5.44×10^6
24	3.08×10^7	

Exercises

6. Show that Equation (9) is a logistic equation. That is, find values of A, B, and k in Equation (8) that make it equivalent to Equation (9).

7. Verify the entries in column 3 of Table 1, and fill in the blanks of that column.

5. DETERMINING MATHEMATICALLY WHETHER GROWTH IS LOGISTIC

Suppose we suspect a certain population growth satisfies a logistic equation. Is there any easy way to determine if this is true, and if so, to determine A, B, and k?

5.1 The Technique

Here is one technique. Look at Equation (7). If we take the natural logarithms of both sides we get

$$\ln \frac{A - f(t)}{f(t)} = \ln Be^{-Akt} \qquad (10)$$

or

(10) $$\ln \frac{A - f(t)}{f(t)} = (-Ak)t + \ln B.$$

The expression on the left side of Equation (10) is a new function of t. Let us give it a name:

(11) $$g(t) = \ln \frac{A - f(t)}{f(t)}.$$

Now if Equation (10) is true (that is, if the population growth is logistic), then we can combine Equations (11) and (10) to get

(12) $$g(t) = (-Ak)t + \ln B.$$

The graph of g(t) is then a straight line with slope -Ak and y-intercept ln B. The experimental points (t,f(t)) will lie near the logistic curve given by Equation (8) if and only if the points (t,g(t)), where g(t) is computed from Equation (11), lie near this straight line.

There is one problem, though. Our new function g(t) is expressed in terms of f(t) and A. The values of f(t) are the experimental data, but what is A? We said in Section 2 that A is a limiting value which the population can never exceed. If we know A beforehand, fine. But quite likely we will not. What then?

If the population growth really is logistic, then it is plain from Figure 3, and can easily be confirmed from Equation (8), that f(t) eventually levels off near a certain value, and that this value is in fact A (the curve approaches A asymptotically, or $\lim_{t\to\infty} f(t) = A$). If we have gathered data for a long enough time then this levelling off should be apparent. In this case we can make a rough plot of the data points and estimate by eye the height of the horizontal line they seem to be approaching. Or, even simpler, we may be able to estimate A directly from the numerical vaues of f(t). If the last few values are nearly equal to each other we can take A to be equal to or just slightly greater than the largest of these values of f(t). Whichever way we do it we can then use this estimate of A to calculate g(t).

5.2 An Example

To illustrate this technique, let us consider the data in Table 2 for a hypothetical population growth. We want to determine whether the growth is logistic and, if it is, to find a reasonable logistic equation to describe it.

TABLE 2

Hypothetical Population Growth

t (hours)	0	2	4	6	8	10	12	14	16
f(t)	40	167	558	1313	2020	2307	2383	2403	2408

We shall begin by plotting the points $(t, f(t))$, as shown in Figure 5. Since we are doing this only to estimate the value of A, we can settle for a very rough sketch. The pattern in Figure 5 looks like Figure 3, which is encouraging. The points appear to level off at about 2410 and we shall use this value as our estimate of A.

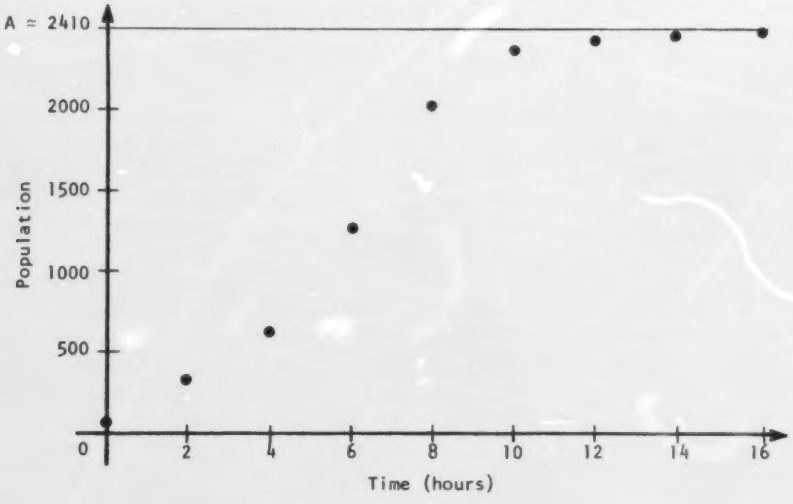


Figure 5. Hypothetical population growth (data from Table 2).

We may now use Equation (11) to calculate the values of g(t). Setting

$$g(t) = \ln \frac{A - f(t)}{f(t)} = \ln \frac{2410 - f(t)}{f(t)},$$

we get the values shown in Table 3.

TABLE 3

Values of g(t) for the Hypothetical Population Growth Described Above

t (hours)	0	2	4	6	8	10	12	14	16
g(t)	4.08	2.60	1.20	-0.18	-1.64	-3.11	-4.64	-5.84	-7.09

The points (t,g(t)) are plotted in Figure 6 They are extremely close to lying on a straight line. It is pretty clear that this growth is logistic.

To find a good logistic equation for this growth we begin by drawing in, by eye, the "best-fitting" straight line on Figure 6. To do this it is wisest to use a transparent straightedge, so it does not block our view of any of the points. We try to draw the line so that some of the points are slightly above it, some are slightly below it, and none are too far away from it.*

It is not important whether the line actually passes through any of the points. In fact, it might well be a mistake simply to draw the line determined by any two of the points. In doing so, we might be unfortunate enough to pick two points which are inaccurate because of experimental error.

*There are a number of mathematical criteria for deciding what lines to use to fit data, among them the criterion of *least squares*, discussed in many calculus texts. In the present unit, however, we can get by with fitting lines by eye in the manner just described.

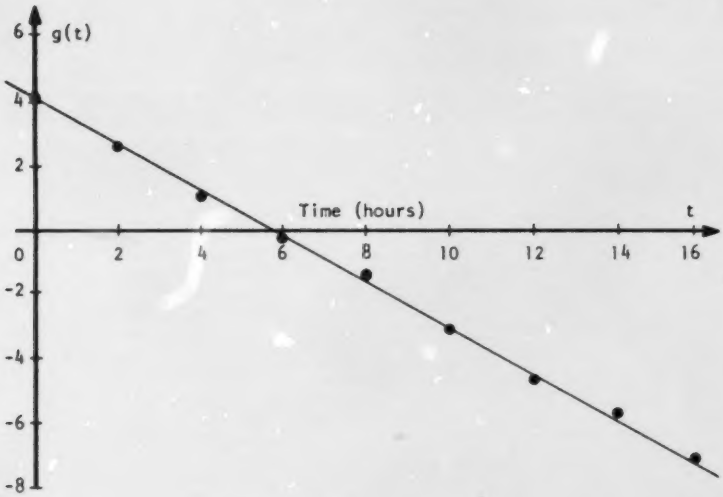


Figure 6. Graph of $g(t) = \ln \frac{A - f(t)}{f(t)}$ for the data in Table 3.

The line we have drawn nearly passes through (0,4) and through (15,-6.6). Therefore its y-intercept is about 4 and its slope is about

$$\frac{-6.6 - 4}{15 - 0} \approx \frac{-10.6}{15} \approx -0.7.$$

According to Equation (12), then,

$$-Ak \approx -0.7$$

and

$$\ln B \approx 4,$$

from which we may estimate

$$B \approx 55.$$

If we use these values for -Ak and B, and recall our estimate that A is about 2410, Equation (8) becomes

$$(13) \qquad f(t) = \frac{2410}{1 + 55e^{-0.7t}}.$$

You may want to calculate the values of f(0), f(2), ... , f(16) from Equation (13), to see how close they come to the observed values in Table 2.

5.3 Three Words of Advice

Notice that in finding the slope of the line in Section 5.2, we used two points on the line itself, rather than two of the points (t,g(t)). In any experiment there are bound to be slight errors, and any two points we select might just happen to be bad ones. The line in effect "averages out" these errors.

Notice also that we chose two points fairly far apart. If we had chosen them close together, then the denominator in our slope formula would have been small, and a slight error in reading the coordinates of either one of the points might have resulted in a huge error in estimating the slope.

Finally, given ln B ≈ 4, we wrote B ≈ 55, although any calculator or book of tables gives $e^4 \approx 54.60$. Remember that the 4 is simply our visual estimate of the coordinate at which the line crosses the vertical axis. It is not a precise reading. The actual value might be as high as 4.05, which would give $B \approx e^{4.05} \approx 57.40$. Besides, the line itself is determined by eye, and might well be higher or lower. So all we can safely say is that B is about equal to 55.

Exercises

8. Here are typical data for the growth of a yeast population:

t (hours)	0	3	6	9	12	15	18
f(t)	9.6	47	175	441	595	651	662

a. Make an estimate of A by graphing f(t) and determining by eye the value at which it appears to level off.

b. Determine whether the growth is logistic.

c. If so, determine -Ak and ln B graphically, and write down the resulting logistic equation.

9. Repeat Exercise 8 for the following data.

t (hours)	0	1	3	6	9	12
f(t)	12	49	100	141	163	170

6. CRITICISM OF THE MODEL

We have already remarked that the mathematical model we have been describing is a primitive one. Perhaps in response to Exercise 3 you have already found some things wrong with it.

The fact that a certain population growth satisfies a logistic equation does not in itself prove that assumptions (a) and (b) of Section 2.1 are correct, nor even that Section 1 fully describes the situation. At best it confirms that these sections do in fact present one possible description. But there are other assumptions which lead mathematically to the logistic equation. And, for any given population data, there are probably other equations which fit the data as well as the logistic equation. Once a mathematician has fit the equation to the data (as in Section 5, for example) it is for the scientist to decide what conclusions may be drawn, possibly with the aid of additional experimental evidence.

Our model is best suited to the growth of unicellular organisms and small multicellular organisms. However, it has been superseded to a great extent by more sophisticated models. If you are interested in a more critical discussion of the model described in this unit, you can find one on pages 30-32 of Pielou (see References). See also the works of Feller and of Krebs.

7. POSTSCRIPT

Most growth curves in the real world have the "sigmoid" shape shown in Figure 1, but in some cases the factors that limit growth stop it precipitously, giving the curve the shape of a J. Although the logistic curve best describes growth of unicellular organisms and small multicellular organisms, logistic curves may be obtained for larger animals under certain conditions.

8. REFERENCES

Feller, W. 1940. On the logistic law of growth and its empirical verification. Acta Biotheoretica (Leiden) 5:51-64.

Howland, J.L., and Grobe, G.A., Jr. 1972. A Mathematical Approach to Biology. Lexington, Mass.: D.C. Heath & Co.

Krebs, C.J. 1972. Ecology. New York: Harper and Row.

Lotka, A.J. 1956. Elements of Mathematical Biology. Dover, New York

Olinick, M. 1978. An Introduction to Mathematical Models in the Social and Life Sciences. Reading, Mass.: Addison-Wesley Publishing Company.

Pielou, E.C. 1969. An Introduction to Mathematical Ecology. New York: Wiley-Interscience.

For bacterial growth curves see:

Davis, B.D., Dulbecco, R., Eisen, H., Ginsberg, H.S., Wood, W.B., Jr., and McCarty, M. 1973. Microbiology. (2nd ed.). New York: Harper and Row.

For populations see:

Wangersky, P.J., and Cunningham, W.J. 1956. On time lags in equations of growth. Proc. Natl. Acad. Sci. 42: 699-702.

Smith, F.E. 1963. Population Dynamics in Daphnia magna and a new model for population growth. Ecology 44: 651-653.

Slobodkin, L.B. 1962. Growth and Regulation of Animal Populations. New York: Holt, Rinehart and Winston.

The Project would like to thank Judith M. Elkins of Sweet Briar College, Sweet Briar, West Virginia for her review, and all others who assisted in the production of this unit.

This unit was field-tested and/or student reviewed at Franklin College, Franklin, Indiana, Essex County College, Newark, New Jersey, Humboldt State University, Arcata, California, Blackburn College, Carlinville, Illinois, and Georgia State University, Atlanta, Georgia, and has been revised on the basis of data received from these sites.

ANSWERS TO EXERCISES

1. Since $f(t) > 0$, we may divide both sides of the differential equation by $f(t)$ to obtain

$$\frac{f'(t)}{f(t)} = k_1.$$

When we integrate both sides of this equation, we have

$$\ln|f(t)| = k_1 t + C,$$

or

$$|f(t)| = e^{k_1 t + C}$$

$$= e^{C} \cdot e^{k_1 t}$$

$$= B_1 e^{k_1 t}.$$

The constant $B_1 = e^{C}$ is positive because e raised to any power is positive. Finally, $|f(t)| = f(t)$ because $f(t) > 0$, and we have

$$f(t) = B_1 e^{k_1 t}, \qquad B_1 > 0.$$

2. Since $f(t) < A$, the quantity $[A-f(t)]$ is positive, and we may divide both sides of the given differential equation by it to obtain

$$\frac{f'(t)}{A - f(t)} = k_2.$$

When we integrate both sides of this equation, we have

$$-\ln|A - f(t)| = k_2 t + C,$$

$$\ln|A - f(t)| = -k_2 t - C,$$

or

$$|A - f(t)| = e^{-k_2 t - C}$$

$$= e^{-C} e^{-k_2 t}$$

$$= B_2 e^{-k_2 t}.$$

The constant $B_2 = e^{-C}$ is positive because it is a power of e. Finally, $|A - f(t)| = A - f(t)$, because $f(t) < A$. This means that

$$A - f(t) = B_2 e^{-k_2 t},$$

or

$$f(t) = A - B_2 e^{-k_2 t}.$$

3. Some considerations to be taken into account are:

   - Length of time between birth and reproductive age;
   - Difficulty of a population surviving if its number drops below a certain value;
   - Changes in reproductive rate because of inbreeding;
   - Sudden changes in the environment;
   - Changes in the survival rate due to changes in organization possible with larger population.

   The equations

$$f'(t) = kf(t)[A-f(t)]^2$$

$$f'(t) = kf(t)e^{-1/[A-f(t)]}$$

   would each satisfy both assumptions (a) and (b).

   The model should work well for simple organisms whose growth rate depends upon their population density, such as bacteria, yeast, and many insects. It may not fit well the growth of social insects, such as bees and ants; animals that can make extensive modifications to their environment or reproductive

cycles, such as humans; or animals whose numbers are so low that they are on the verge of extinction, such as whooping cranes and ivory-billed woodpeckers.

4. a. $t = \frac{1}{Ak} \ln 3$; $f(t) = \frac{A}{2}$.

   b. $\left(\frac{1}{Ak} \ln 8, \frac{A}{2}\right)$.

5. a. 500

   b. $\frac{500}{1+e^2} \approx 59.6$.

   c. 20 days.

6. $A = 3.1 \times 10^7$; $B = 22026$; $k = 2.02 \times 10^{-8}$.

7. 2.63×10^3; 4.40×10^4; 7.21×10^5; 3.08×10^7.

8. a. A is about 665. (Answers will vary.)

   b. Yes.

   c. $-Ak \approx -0.55$; $\ln B \approx 4.22$; $f(t) \approx \frac{665}{1+68.03e^{-0.55t}}$.

9. a. $A \approx 178$. (Answers will vary.)

   b. No.

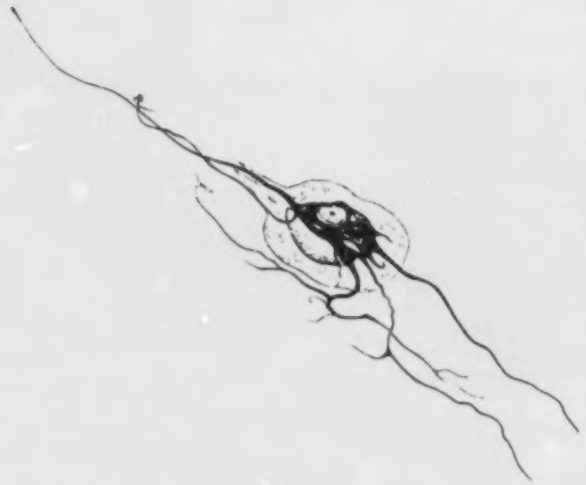
umap

UNIT 70

MODULES AND MONOGRAPHS IN UNDERGRADUATE MATHEMATICS AND ITS APPLICATIONS PROJECT

SELECTION IN GENETICS

by Brindell Horelick and Sinan Koont



BIOLOGICAL APPLICATIONS OF CALCULUS

UNITS 67-70

edc/umap /55chapel st./newton, mass. 02160

SE 036 474

SELECTION IN GENETICS

Brindell Horelick and Sinan Koont
Department of Mathematics
University of Maryland Baltimore County
Baltimore, Maryland 21228

8/11/77

TABLE OF CONTENTS

1. THE DISAPPEARRANCE OF RECESSIVE GENES 1
1.1 Introduction 1
1.2 Gene Frequency 1
1.3 Selective Disadvantage 2

2. REVIEW OF GENETICS 2
2.1 The Hardy-Weinberg Ratio 2
2.2 Calculation of Gene Frequencies 3

3. RECURRENCE RELATION FOR GENE FREQUENCY 3
3.1 Introduction 3
3.2 The N-th Generation Gene Distribution 4
3.3 The Recurrence Relation 4
3.4 Use of the Recurrence Relation 5
3.5 Computational Results 5
3.6 Disadvantage of Recurrence Relation 6

4. A COMPUTER-FREE PROCESS 6
4.1 A Differential Approximation 6
4.2 The Approximation Formula 9
4.3 Comparison of Results 9

5. ANSWERS TO EXERCISES 10

Intermodular Description Sheet: *UMAP Unit 70*

Title: SELECTION IN GENETICS

Authors: Brindell Horelick and Sinan Koont
Department of Mathematics
University of Maryland Baltimore County
Baltimore, Maryland 21228

Review Stage/Date: III 8/11/77; II 5/26/77; I 10/31/76

Classification: BIO APPL CALC / SELC IN GENETICS (U70)

Suggested Support Materials:

References:

Haldane, J. B. S. (1932). The Causes of Evolution. Longman, Green and Co., London.

Lotka, A. J. (1956). Elements of Mathematical Biology. Dover, New York.

Smith, J. M. (1968). Mathematical Ideas in Biology. Cambridge, Cambridge.

Wilson, E. O. and Bossert, W. H. (1971). A Primer of Population Biology. Sinaure Associates, Inc., Stamford, Connecticut.

Prerequisite Skills:

1. Know the formula for the slope of a line segment.
2. Know that f'(x) describes a tangent slope.
3. Be able to integrate $\int \frac{f'(x)}{f(x)}\,dx$.

This unit is intended for Calculus students with an active interest in, and some background knowledge of, biology. This background knowledge may be represented by concurrent registration in a college level introductory biology course.

Output Skills:

1. Be able to describe how a reccurence relation for gene frequency in the n-th generation is obtained.
2. Be able to explain the use of calculus to approximate the result obtained from this reccurence relation.

Other Related Units:

The following units involve the exponential function and apply simple differential equations to biology.

Reaction Time and the Central Nervous System *(Unit 67)*
Population Growth and the Logistic Curve *(Unit 68)*
The Digestive Process of Sheep *(Unit 69)*

© 1977 EDC/Project UMAP
All Rights Reserved.

MODULES AND MONOGRAPHS IN UNDERGRADUATE
MATHEMATICS AND ITS APPLICATIONS PROJECT (UMAP)

The goal of UMAP is to develop, through a community of users and developers, a system of instructional modules in undergraduate mathematics which may be used to supplement existing courses and from which complete courses may eventually be built.

The Project is guided by a National Steering Committee of mathematicians, scientists, and educators. UMAP is one of many projects of Education Development Center, Inc., a publically supported, nonprofit corporation engaged in educational research and development in the U.S. and abroad.

PROJECT STAFF

William U. Walton	Senior Pedagogical and Editorial Advisor
Ross L. Finney	Senior Mathematics Editor
Solomon Garfunkel	Consortium Coordinator
Felicia Weitzel	Associate Director for Administration
Barbara Kelczewski	Editorial/Production Assistant
Dianne Lally	Project Secretary
Paula Santillo	Financial Assistant/Secretary

NATIONAL STEERING COMMITTEE

W.T. Martin	MIT (Chairman)
Steven J. Brams	New York University
Llayron Clarkson	Texas Southern University
James D. Forman	Rochester Institute of Technology
Ernest J. Henley	University of Houston
William F. Lucas	Cornell University
Walter E. Sears	University of Michigan Press
Peter Signell	Michigan State University
George Springer	Indiana University
Robert H. Tamarin	Boston University
Alfred B. Willcox	Mathematical Association of America
Nancy J. Kopell	Northeastern University

The Project would like to thank Nathan F. Simms, Jr., Charles Votaw, and Edwin H. Rogers for their reviews, and all others who assisted in the production of this unit.

This material was prepared with the support of National Science Foundation Grant No. SED 76-19615. Recommendations expressed are those of the authors and do not necessarily reflect the views of the NSF, nor of the National Steering Committee.

1. THE DISAPPEARANCE OF RECESSIVE GENES

1.1 Introduction

Suppose we have a population of a certain organism with two *allelic* genes *A* and *a*, with *a* being the recessive gene. In such a population there are three genotypes: the homozygotes *AA* and *aa* and the heterozygote *Aa*. Suppose the homozygotes of the recessive gene have less chance of surviving to adulthood than the other two zygotes. This appears to be the case, at least in the United States, for the recessive gene (hemoglobin S) associated with sickle cell anemia, a disease found primarily among black people.[1] What can we say about the occurrence of *a* and *A* with the passage of generations? It seems likely that *aa* will eventually be 'bred out' of the population, but how fast?

1.2 Gene Frequency

To word this question more precisely, let us review some genetic terminology. The quantitative measure of the occurrence of a gene is its *gene frequency*. The *gene frequency*, p, of a gene, *A*, is defined as the ratio of the number of genes of type *A* to the total number of allelic genes present:

$$(1) \qquad p = \frac{\#\text{ of } A\text{-genes}}{\#\text{ of } A\text{-genes} + \#\text{ of } a\text{-genes}} = \frac{\#\text{ of } A\text{-genes}}{2N},$$

where N is the total number of organisms in the population.

---

[1] The authors wish to thank William U. Walton of Education Development Center for calling this observation to their attention.

1.3 Selective Disadvantage

If a certain fraction, s, of the zygotes *Aa* and *AA* survive to adulthood, while only a fraction, (1 - k)s, of the zygote *aa* does (where $0 < k < 1$), then the zygote *aa* is said to have a *selective disadvantage*, k. Since $(1 - k)s < s$, the zygote *aa* is less "fit" than *Aa* or *AA*. This disadvantage is often expressed as a percentage. For example, a selective disadvantage of 1% means $k = 0.01$.

Let us denote the respective frequencies of *A* and *a* genes in the n-th generation by p_n and q_n. With this terminology, we can state our original question as follows: given gene frequencies p_0 and q_0 initially, given that *aa* has a selective disadvantage k, and given any positive $q < q_0$, how many generations will pass before $q_n < q$?

Before developing our mathematical model, let us review some more genetic terminology and some genetic theory.

2. REVIEW OF GENETICS

2.1 The Hardy-Weinberg Ratio

Consider an adult population with allelic genes *A* and *a* having gene frequencies p and q respectively. Suppose it mates randomly (that is, the probability of any one organism mating with an *AA*, *Aa* or *aa* genotype does not depend on its own genotype). The genotypes *AA*, *Aa* and *aa* will occur in the offspring population in the proportion

$$(2) \qquad p^2 : 2pq : q^2,$$

called the Hardy-Weinberg Ratio. To see this, think of forming a zygote by choosing one gene at random and then choosing the other at random. The fraction of first choices which will yield an *A* gene is p. The fraction of *these* for which the second choice is also an *A* gene is p.

So altogether a fraction p^2 of "pair-choices" will result in *AA*. Similarly, a fraction q^2 of the choices will result in *aa*, leaving $1 - p^2 - q^2 = 2pq$ for *Aa*.

2.2 Calculation of Gene Frequencies

If we know the genotypes *AA*, *Aa* and *aa* occur in the proportion P : Q : R, we can compute the gene frequencies p and q. This is particularly easy if P + Q + R = 1, for then P, Q and R are the actual fractions of the population belonging to each genotype. If there are N organisms, there are PN of type *AA* (each having two *A* genes) and QN of type *Aa* (each having one *A* gene), so there are (2PN + QN) *A* genes altogether. Therefore, using equation (1),

$$(3) \qquad p = \frac{2PN + QN}{2N} = P + \frac{Q}{2}.$$

Since p + q = 1,

$$(4) \qquad q = 1 - p = 1 - P - \frac{Q}{2} = P + Q + R - P - \frac{Q}{2}$$

$$= R + \frac{Q}{2}.$$

3. A RECURRENCE RELATION FOR GENE FREQUENCY

3.1 Introduction

We can now develop our mathematical model. We assume there is no mating between generations, and that within each generation the mating is random. We shall compute the gene frequencies for any generation, using a *recurrence relation*. This means we will obtain an expression giving the gene frequencies for any generation in terms of those for the preceding generation. We will then apply the formula n times, working our way up one generation at a time from generation zero to generation n.

3.2 The N-th Generation Gene Distribution

Since the gene frequencies in the n-th generation are p_n and q_n (where $q_n = 1 - p_n$), by formula (2) the proportion of genotypes *AA*, *Aa* and *aa*, in the (n + 1)-th generation will be:

$$p_n^2 : 2p_nq_n : q_n^2 .$$

However, because of the different survival rates we have assumed, the proportion in the (n + 1)-th adult generation changes to:

$$p_n^2 : 2p_nq_n : (1 - k)q_n^2 .$$

3.3 The Recurrence Relation

The preceding proportion can be rewritten:

$$hp_n^2 : 2hp_nq_n : h(1 - k)q_n^2$$

for any positive h.

Let us choose h so that $hp_n^2 + hp_nq_n + h(1 - k)q_n^2 = 1$, for then we can apply equations (3) and (4) with $P = hp_n^2$, $Q = 2hp_nq_n$, and $R = h(1 - k)q_n^2$, getting:

$$(5) \qquad p_{n+1} = hp_n^2 + 2hp_nq_n$$

$$(6) \qquad q_{n+1} = h(1 - k)q_n^2 + hp_nq_n .$$

We can eliminate the bothersome quantity h by dividing equation (5) by equation (6).

$$(7) \qquad \frac{p_{n+1}}{q_{n+1}} = \frac{p_n^2 + p_nq_n}{(1 - k)q_n^2 + p_nq_n} = \frac{\left(\frac{p_n}{q_n}\right)^2 + \left(\frac{p_n}{q_n}\right)}{(1 - k) + \left(\frac{p_n}{q_n}\right)}$$

In this last step, we have simply multiplied the numerator and denominator of our expression by $1/q_n$. As you see, the *ratio* p_n/q_n of gene frequencies plays a

central role in equation (7). If we call this ratio y_n, we can rewrite equation (7):

$$(8) \qquad y_{n+1} = \frac{y_n^2 + y_n}{(1 - k) + y_n} = \frac{y_n(y_n + 1)}{y_n + 1 - k},$$

the recurrence relation we seek.

3.4 Use of the Recurrence Relation

Starting from any given q and k, we can compute

$$y_0 = \frac{p_0}{q_0} = \frac{1 - q_0}{q_0},$$

and then use equation (8) to evaluate successively y_1, then y_2, then y_3, etc. Having computed y_n, we recall that

$$y_n = \frac{p_n}{q_n} = \frac{1 - q_n}{q_n},$$

from which we obtain

$$q_n = \frac{1}{1 + y_n}.$$

We use this equation to compute q_n. We then compare q_n with q, and repeat the entire process until $q_n < q$.

3.5 Computational Results

This procedure was programmed on a computer for a recessive gene with initial gene frequency q = 0.99 and selective disadvantage of 1% (k = 0.01).[2] The results are given in Table 1.

Table 1

Gradual Disappearance of Recessive Gene

q_n	$y_n = \frac{1 - q_n}{q_n}$	n(from equation (8))
0.99	0.01	0
0.1	9.0	1,574
0.01	99.0	10,813
0.001	999.0	101,125

[2] Computer work supported by the University of Maryland Computer Science Center.

We see that at first the *a* gene declines in frequency quite rapidly. It goes down from 99% to 10% in 1574 generations; but it takes nearly 100,000 generations to go down from 10% to 0.1%. The persistence of the gene can be explained qualitatively by noting that as the frequency declines, the few remaining *a* genes are almost all in heterozygotes *Aa* and therefore not at a selective disadvantage.

3.6 Disadvantage of Recurrence Relation

The procedure just described is heavily computer-dependent. Without a computer it is at best extremely cumbersome and probably not even feasible. Let us see how we can use calculus to develop a computer-free procedure for obtaining excellent approximations to the results in Table 1.

4. A COMPUTER-FREE PROCESS

4.1 A Differential Approximation

We shall attempt to find a *differentiable* function f, defined for all $x > 0$, with the property that whenever x equals an integer n we have:

$$(9) \qquad f(n) = y_n.$$

From equation (8) we can compute:

$$(10) \qquad y_{n+1} - y_n = \frac{y_n(y_n + 1)}{y_n + 1 - k} - y_n = \frac{k y_n}{y_n + 1 - k}.$$

Therefore for equation (9) to be true we must have:

$$(11) \qquad f(n + 1) - f(n) = k \frac{f(n)}{f(n) + 1 - k}.$$

Let us see what happens if we take equation (11), which is written for *integer* values of the variable, and require it to hold for all $x \geq 0$:

$$(12) \qquad f(x + 1) - f(x) = k \frac{f(x)}{f(x) + 1 - k}.$$

The left side of equation (12) can be written as

$$\frac{f(x+1) - f(x)}{1}$$

or

$$\frac{f(x+1) - f(x)}{(x+1) - x}.$$

In other words, it is the slope of the line segment between the points S(x, f(x)) and T(x + 1, f(x + 1)) on the graph of f.

Now equations (9) and (10) tell us that, for integer values of n,

$$f(n+1) - f(n) \approx y_{n+1} - y_n < k.$$

Therefore, *for small* k, f(n + 1) - f(n) is small. (Notice, for example, that in Table 1 it takes 1574 generations for y_n to increase by 8.99.) This means (see Figure 1) that a rather flat curve, representing our function f, can be drawn satisfying equation (9). For such a curve (see Figure 2) the slope of the segment $\overline{ST}$ is nearly equal to the tangent slope at S, which is of course f'(x).

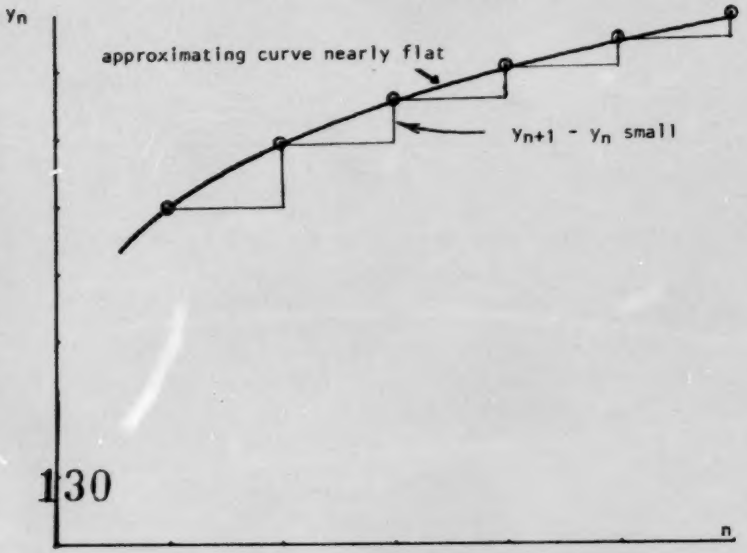


Figure 1. Hypothetical Values of y_n for small k, and Hypothetical Approximating Curve

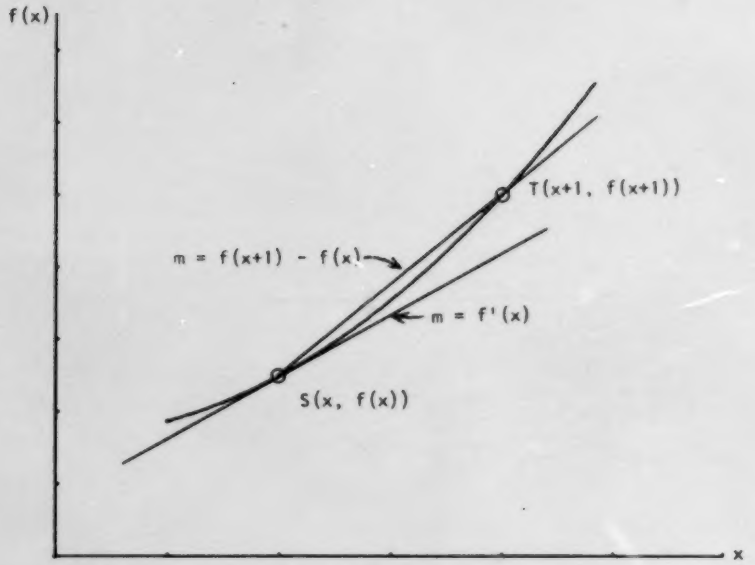


Figure 2. Tangent Slope and Segment Slope for Nearly Flat Curve (Figure Exaggerated for Clarity).

In view of these remarks, it seems reasonable to replace equation (12) by:

$$(13) \qquad f'(x) = k \frac{f(x)}{f(x) + 1 - k}.$$

Exercise 1: Show that any funciton f which satisfies equation (13) for small positive k does indeed satisfy equation (11).

HINTS: (a) For $n \le x < n+1$, explain why $f(x) \approx f(n)$ and therefore

$$(14) \quad f'(x) \approx k \frac{f(n)}{f(n) + 1 - k}.$$

(b) Integrate equation (14) from n to n + 1, noticing that the right side is a constant.

Exercise 2: The converse of Exercise 1 is false. Even for very small k, there are differentiable functions satisfying equation (12) but *not* equation (13). Find and example of one.

HINT: You can even let k = 0, and consider something like the sine function.

4.2 The Approximation Formula

Let us now try to solve equation (13) for the function f.

$$f'(x)\,\frac{f(x) + 1 - k}{f(x)} = k$$

$$f'(x) + (1-k)\,\frac{f'(x)}{f(x)} = k$$

$$\int f'(x)dx + (1-k)\int \frac{f'(x)}{f(x)}\,dx = \int k\,dx$$

(15) $$f(x) + (1-k)\ln f(x) = kx + C.$$

Putting $x = 0$ and solving for C we obtain

$$C = f(0) + (1-k)\ln f(0),$$

and so equation (15) becomes:

(16) $$f(x) - f(0) + (1-k)\ln\frac{f(x)}{f(0)} = kx.$$

Going back to equation (9) we get

$$y_n - y_0 + (1-k)\ln\frac{y_n}{y_0} \simeq kn.$$

Solving for n we find

(17) $$n \simeq \frac{1}{k}\left[y_n - y_0 + (1-k)\ln\frac{y_n}{y_0}\right].$$

Exercise 3: Use equation (16) to prove:

$$f(x)^{1-k} = He^{kx-f(x)}, \text{ where } H = e^{f(0)}\,f(0)^{1-k}.$$

4.3 Comparison of Results

Equation (17) provides the computer-free procedure we are seeking. Instead of using equation (8) over 100,000 times to complete the last row of Table 1, for example, we need only use equation (17) once. The calculation can easily be made with a hand calculator, or with a pencil and paper and a table of logarithms (base e). In every case in Table 1, the result is well within one-half of one percent of the true n. (See Exercise 4.)

Exercise 4: Use equation (17) to approximate the values of n in Table 1. In each case express the error as a percentage of the value given in the table.

Exercise 5: Use equation (17) to find approximate values of n
(a) if $k = 0.02$
(b) if $k = 0.05$.
In each case, take $q_0 = 0.99$, and let q_n have each of the values in Table 1.

Exercise 6: Suppose that homozygotes of a recessive gene have a 3% *greater* chance of survival then either *AA* or *Aa*. The initial frequency of *a* is 0.002. Find an approximate value for the number of generations needed before the gene frequency of *a* reaches 60%.

---

5. ANSWERS TO EXERCISES

1. (a) If k is small (remember that it already lies between 0 and 1) then the number

$$k\,\frac{f(x)}{f(x) + 1 - k}$$

in equation (12) is a small positive number because it is positive and less than k. Under these circumstances, equation (12) tells us that $f(x+1) \simeq f(x)$, and the value of f(x) increases very little (in fact by less than k) as x goes from n to n+1. Therefore, $f(x) \simeq f(n)$ for $n \le x < n+1$. We obtain equation (14) when f(x) is replaced by f(n) in Equation (13).

(b) $$\int f'(x)\,dx \simeq \int_n^{n+1} k\,\frac{f(n)}{f(n) + 1 - k}\,dx$$

$$f(x) \simeq k\,\frac{f(n)}{f(n) + 1 - k}\,x\,\Big|_{x=n}^{x=n+1}$$

$$\simeq k\,\frac{f(n)}{f(n) + 1 - k}.$$

2. One example is $f(x) = \frac{1}{2} \sin 2\pi x$. With $k = 0$, the right sides of equations (12) and (13) are both zero. The left side of (12) is zero for all x. The left side of (13), however, is $\pi \cos 2\pi x \neq 0$.

3. Selected steps in the derivation:

$$e^{f(x)} \cdot e^{-f(0)} \cdot e^{(1-k)\ln f(x)/f(0)} = e^{kx}$$

$$\left(e^{\ln f(x)/f(0)}\right)^{1-k} = e^{kx - f(x) + f(0)}$$

$$\left(\frac{f(x)}{f(0)}\right)^{1-k} = e^{kx - f(x)} \cdot e^{f(0)}$$

$$f(x)^{1-k} = e^{kx - f(x)} \cdot e^{f(0)} \cdot f(0)^{k-1}$$

$$= He^{kx - f(x)}.$$

4.

q_n	n	%error
0.1	1,572	0.13
0.01	10,809	0.04
0.001	101,038	0.09

5.

q_n	n (for k = 0.02)	n (for k = 0.05)
0.1	783	309
0.01	5,400	2,155
0.001	50,514	20,199

6. 16,838

STUDENT FORM 1

Request for Help

Return to:
EDC/UMAP
55 Chapel St.
Newton, MA 02160

Student: If you have trouble with a specific part of this unit, please fill out this form and take it to your instructor for assistance. The information you give will help the author to revise the unit.

Your Name_____ Unit No._____

Page_____
○ Upper
○ Middle
○ Lower

OR

Section_____
Paragraph_____

OR

Model Exam
Problem No._____
Text
Problem No._____

Description of Difficulty: (Please be specific)

Instructor: Please indicate your resolution of the difficulty in this box.

○ Corrected errors in materials. List corrections here:

○ Gave student better explanation, example, or procedure than in unit. Give brief outline of your addition here:

○ Assisted student in acquiring general learning and problem-solving skills (not using examples from this unit.)

Instructor's Signature_____

Please use reverse if necessary.

STUDENT FORM 2

Unit Questionnaire

Return to:
EDC/UMAP
55 Chapel St.
Newton, MA 02160

Name_____Unit No._____Date_____

Institution_____Course No._____

Check the choice for each question that comes closest to your personal opinion.

1. How useful was the amount of detail in the unit?

____Not enough detail to understand the unit
____Unit would have been clearer with more detail
____Appropriate amount of detail
____Unit was occasionally too detailed, but this was not distracting
____Too much detail; I was often distracted

2. How helpful were the problem answers?

____Sample solutions were too brief; I could not do the intermediate steps
____Sufficient information was given to solve the problems
____Sample solutions were too detailed; I didn't need them

3. Except for fulfilling the prerequisites, how much did you use other sources (for example, instructor, friends, or other books) in order to understand the unit?

____A Lot ____Somewhat ____A Little ____Not at all

4. How long was this unit in comparison to the amount of time you generally spend on a lesson (lecture and homework assignment) in a typical math or science course?

Much	Somewhat	About	Somewhat	Much
____Longer	____Longer	____the Same	____Shorter	____Shorter

5. Were any of the following parts of the unit confusing or distracting? (Check as many as apply.)

____Prerequisites
____Statement of skills and concepts (objectives)
____Paragraph headings
____Examples
____Special Assistance Supplement (if present)
____Other, please explain_____

6. Were any of the following parts of the unit particularly helpful? (Check as many as apply.)

____Prerequisites
____Statement of skills and concepts (objectives)
____Examples
____Problems
____Paragraph headings
____Table of Contents
____Special Assistance Supplement (if present)
____Other, please explain_____

Please describe anything in the unit that you did not particularly like.

Please describe anything that you found particularly helpful. (Please use the back of this sheet if you need more space.)